MASZSTABINVARIANTE KORRELATIONSTHEORIE

INAUGURALDISSERTATION
ZUR ERLANGUNG DES DOKTORGRADES
GENEHMIGT VON DER
MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT
DER FRIEDRICH-WILHELMS-UNIVERSITÄT
ZU BERLIN

VON

WASSILIJ HÖFFDING
AUS ST. PETERSBURG (RUSSLAND)

TAG DER MÜNDLICHEN PRÜFUNG: 24. JULI 1940
TAG DER PROMOTION: 14. DEZEMBER 1940

DRUCK VON B. G. TEUBNER IN LEIPZIG

N 64

REFERENTEN:
PROFESSOR DR. A. KLOSE
PROFESSOR DR. H. GEPPERT

R. UNIV.
BIBLIOTHEEK
LEIDEN

Sonderabdruck aus den „Schriften des Mathematischen Instituts und des Instituts für angewandte Mathematik der Universität Berlin“ / Band 5

Einleitung.

Als die Aufgabe der Korrelationsrechnung kann man die Untersuchung derjenigen Eigenschaften der mehrdimensionalen Verteilungen bezeichnen, die für diese Verteilungen charakteristisch sind, d. h. bei eindimensionalen Verteilungen nicht auftreten. Diese Eigenschaften hängen vor allem mit den Beziehungen der Merkmale untereinander zusammen. Aus der Gesamtheit dieser zum Gegenstand der Korrelationsrechnung gehörenden Eigenschaften soll im folgenden eine besondere Klasse von Eigenschaften näher betrachtet werden.

Bei der Bearbeitung eines statistischen Materials, bei dem *quantitative* Merkmale (zufällige Variable) $\xi, \eta, \ldots$ vorliegen, kann es sich als zweckmäßig erweisen, *den ursprünglichen Maßstab der Merkmale zu ändern.* Eine solche Maßstabänderung kann erforderlich werden unabhängig davon, ob die Merkmale durch Messung oder durch Zählung ermittelt wurden. Handelt es sich um meßbare Merkmale, so wird man unter Umständen zu einer anderen Maßeinheit (etwa von Millimeter zu Zentimeter) übergehen. Bei beliebigen quantitativen — durch Zählung oder Messung ermittelten — Merkmalen kann es von Vorteil sein, in einer oder mehreren der Variablen $\xi, \eta, \ldots$ vom ursprünglichen etwa zum logarithmischen Maßstab überzugehen. Allgemein wird eine solche *Maßstabänderung* durch eine Transformation

$$(M) \qquad \bar{\xi} = f(\xi), \quad \bar{\eta} = g(\eta), \ldots$$

ausgedrückt, wobei wir voraussetzen wollen, daß die Transformation (M)

1. umkehrbar eindeutig (d. h. eindeutig und eindeutig umkehrbar) ist,

2. die Anordnung der Merkmale ungeändert läßt, so daß z. B. aus $\xi_1 < \xi_2$ und $\bar{\xi}_1 = f(\xi_1)$, $\bar{\xi}_2 = f(\xi_2)$ stets $\bar{\xi}_1 < \bar{\xi}_2$ folgt.

Sind die Merkmale $\xi, \eta, \ldots$ kontinuierlich und sind $f(\xi)$ und $g(\eta)$ stetig und stückweise differenzierbar — wie wir im folgenden annehmen werden —, so sind diese beiden Bedingungen gleichbedeutend damit, daß die Ableitungen $f'(\xi)$, $g'(\eta)$, ... endlich, von Null verschieden und positiv sein sollen.[1])

1) Vgl. H. Bruns, *Wahrscheinlichkeitsrechnung und Kollektivmaßlehre.* Berlin 1906, S. 126.

Wenn wir künftig schlechtweg von *Maßstabänderungen* sprechen werden, so werden wir darunter Transformationen der Form (M) verstehen, die die Voraussetzungen 1. und 2. erfüllen.

Geht man nun von den ursprünglichen Merkmalen $\xi, \eta, \ldots$ durch Maßstabänderung zu den neuen Merkmalen $\bar{\xi}, \bar{\eta}, \ldots$ über, so wird die Verteilung der neuen Merkmale eine andere sein, als die der ursprünglichen. Von den Eigenschaften der ursprünglichen Verteilung werden einige auch in den neuen Variablen erhalten bleiben, während andere durch die Maßstabänderung beeinflußt werden. *Man kann demgemäß die Gesamtheit der zum Gegenstand der Korrelationsrechnung gehörenden Eigenschaften einer mehrdimensionalen Verteilung in zwei Klassen einteilen, je nachdem ob sie gegenüber einer beliebigen Maßstabänderung invariant sind oder nicht.*

So ist z. B. der Verlauf der Regressionslinien vom jeweiligen Maßstab abhängig. Auch die mit diesen zusammenhängende Eigenschaft der Nicht-Korreliertheit[2]) ist an den Maßstab gebunden. Beispielsweise ist in der folgenden Wahrscheinlichkeitsverteilung weder ξ mit η noch η mit ξ korreliert.

η \ ξ	-1	0	1
-1	0	$\frac{1}{5}$	0
0	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
1	0	$\frac{1}{5}$	0

Ersetzt man aber ξ durch $\bar{\xi}$, indem man für $\xi = 1$ den Wert $\bar{\xi} = 2$ setzt, während sonst $\bar{\xi} = \xi$ bleibt, so wird der bedingte Erwartungswert von $\bar{\xi}$ nicht mehr für alle η derselbe sein.

Dagegen ist die *stochastische Unabhängigkeit* der Merkmale eine *maßstabinvariante* Eigenschaft. Denn wenn die Merkmale ξ und η voneinander unabhängig sind, so ist die Wahrscheinlichkeit der Merkmalkombination (ξ, η) gleich dem Produkt der Wahrscheinlichkeiten der Einzelmerkmale ξ und η für sich genommen, und bei einer Transformation (M), die sich ja nur auf die Einzelmerkmale erstreckt, bleibt die Produktdarstellung erhalten (wie im übrigen unten in § 1 näher ausgeführt werden wird).

2) Im Sinne von A. Tschuprow, *Grundbegriffe und Grundprobleme der Korrelationstheorie*. Berlin 1925, S. 33. „η ist nicht korreliert mit ξ" bedeutet geometrisch, daß die Regressionslinie von η auf ξ eine zur ξ-Achse parallele Gerade ist.

Ebenso wie die stochastische Unabhängigkeit wird man den „*Grad der stochastischen Abhängigkeit*" zweier Merkmale als eine maßstabinvariante Eigenschaft ansprechen. Denn es dürfte der Vorstellung, die man sich von diesem Begriff macht, entsprechen, daß die durch eine Maßstabänderung (M) eingeführten Merkmale $\bar{\xi}, \bar{\eta}$ weder „stärker" noch „schwächer" stochastisch verbunden sind als die ursprünglichen Merkmale ξ, η. (Dagegen wird man den durch eine Transformation der Form $\bar{\xi} = f(\xi, \eta)$, $\bar{\eta} = g(\xi, \eta)$ eingeführten Merkmalen im allgemeinen einen ganz anderen „Abhängigkeitsgrad" zusprechen als den ursprünglichen Merkmalen.) Dementsprechend wird man von einer Größe, die als *Maß der stochastischen Abhängigkeit* der Merkmale dienen soll, verlangen, daß sie *gegenüber Maßstabänderungen invariant* ist.

Eine solche Forderung wird man zunächst an ein eigentliches Abhängigkeitsmaß stellen, d. h. eine Größe, die nach Art von K. Pearsons mittlerer quadratischer Kontingenz (mean square contingency) einen bestimmten Wert, etwa Null, dann und nur dann annimmt, wenn die Merkmale stochastisch unabhängig sind. Aber auch eine Maßzahl, aus deren Verhalten man zwar nicht die stochastische Unabhängigkeit folgern kann, die aber dafür zwischen „positiver" und „negativer" Verbundenheit der Variablen zu unterscheiden vermag, wie etwa der Korrelationskoeffizient, wird unseres Erachtens im allgemeinen von größerem Wert sein, wenn sie gegenüber Maßstabänderungen invariant ist.

Es sollen hier einige weitere Gesichtspunkte kurz zusammengestellt werden, die zeigen, daß es wünschenswert ist, Begriffe wie „Grad der Abhängigkeit" oder „Grad der Verbundenheit" zweier oder mehrerer zufälliger Variablen als maßstabinvariante Größen zu definieren.

1. Der jeweils benutzte Maßstab kann mehr oder weniger „zufällig" sein, da eine und dieselbe Größe unter Umständen mit verschiedenen Maßeinheiten und auf verschiedene Art und Weise gemessen werden kann.[3])

2. Sind zwei zufällige Variable, deren Zusammenhang untersucht werden soll, in verschiedenen Maßstäben gemessen, so erscheint es angebracht, bei der Ermittlung des Abhängigkeitsgrades dieser Variablen von ihren Maßstäben abzusehen, da sie erst dadurch eigentlich *vergleichbar* werden.[4])

3) Vgl. a) C. Spearman, *The Proof and Measurement of Association between Two Things*, Amer. Journ. of Psychology 15 (1904), S. 79—82; b) Ders., *Footrule for Measuring Correlation*, Brit. Journ. of Psychology 2 (1906), § 2.

4) Vgl. C. Spearman, a. a. O. [3b]), § 2.

3. Auch wird erst ein maßstabinvariantes Korrelations- oder Abhängigkeitsmaß gestatten, den Abhängigkeitsgrad zweier mehrdimensionaler Verteilungen, bei denen die Variablen in verschiedenen Maßstäben gemessen sind, in einwandfreier Weise miteinander zu vergleichen.

Verschiedene in Gebrauch befindliche oder vorgeschlagene Korrelations- und Abhängigkeitsmaße besitzen die Eigenschaft, gegenüber Maßstabänderungen invariant zu sein. Vor allem wären hier PEARSONS mittlere quadratische Kontingenz[5]) und die daraus gebildeten Größen zu nennen, sowie alle Maßzahlen, die an Stelle der ursprünglichen Variablen die Ränge der Merkmale benutzen; zu diesen gehört in erster Linie der von SPEARMAN[6]) eingeführte Rang-Korrelationskoeffizient sowie das von LINDEBERG[7]) vorgeschlagene Korrelationsprozent.[8])

Es verdient hervorgehoben zu werden, daß SPEARMAN die selbständige Bedeutung der „Rangmethode" und den in der Maßstabinvarianz liegenden Vorteil klar erkannt und ausgesprochen hat. Auch LINDEBERG äußert sich in diesem Sinne, wenn er auch sein Korrelationsprozent anscheinend in erster Linie auf Fälle anwenden will, wo die Variablen nicht exakt meßbar sind, sondern nur nach ihrer relativen Größe (nach Rängen) geordnet werden können. Dagegen sah K. PEARSON[9]) in der Ersetzung der ursprünglichen Variablen durch Ränge einen „Rückschritt", und den SPEARMANschen Rang-Korrelationskoeffizienten wollte er nicht als „Ding an sich" gelten lassen, sondern nur als Hilfsmittel zur bequemeren Berechnung des gewöhnlichen Korrelationskoeffizienten r. HOTELLING und PABST[10]) befürworten zwar die Anwendung des Rang-Korrelationskoeffizienten als „test of significance" auf beliebige zweidimensionale Verteilungen, äußern

5a) K. Pearson, *On the Theory of Contingency and its Relation to Association and Normal Correlation*, Drapers' Company Research Memoirs, Biometric Ser. 1, London 1904. — b) Zur Maßstabinvarianz der mittleren quadratischen Kontingenz vgl. R. v. Mises, *Wahrscheinlichkeitsrechnung*. Leipzig 1931, S. 360.

6) A. a. O. [3b]), S. 98.

7) J. W. Lindeberg, *Über die Korrelation*, Den 6. skandinaviske Matematikerkongres i Köbenhavn, 1925, S. 437—446.

8) Offenbar ohne Lindebergs Arbeit zu kennen, hat vor kurzem M. G. Kendall, *A New Measure of Rank Correlation*, Biometrika 30 (1938), S. 81—93, ein Korrelationsmaß vorgeschlagen, das im wesentlichen mit dem Korrelationsprozent übereinstimmt.

9) K. Pearson, *On Further Methods of Determining Correlation*, Drapers' Company Research Memoirs, Biometric Ser. 4, London 1907, S. 9, 12, 32 u. a.

10) H. Hotelling and M. R. Pabst, *Rank Correlation and Tests of Significance Involving No Assumption of Normality*, Ann. Math. Stat. 7 (1936), S. 29—43.

aber, ähnlich wie K. PEARSON, die Ansicht, der Grad der Verbundenheit (degree of relationship) könne ohne spezielle Annahmen, wie etwa die der zweidimensionalen Normalverteilung, nicht gemessen werden. Die Verfasser sagen u. a.: „Ein jedes Maß des Grades der Verbundenheit wird seinen Wert ändern, es sei denn, daß dieser Wert der Unabhängigkeit entspricht, wenn andere Transformationen als die einer begrenzten Klasse auf eine der Variablen angewandt werden.“[11]) Uns scheinen diese Ausführungen wenig einleuchtend. Begriffe wie „Grad der Verbundenheit“ oder „Grad der Abhängigkeit“ können nur durch eine zweckmäßige Definition festgelegt werden. Es wird aber kaum ein Zweifel bestehen, daß es zweckentsprechender ist, solche Begriffe in einer Weise zu definieren, daß sie von Maßstabänderungen nicht abhängen. Daß solche Definitionen sehr wohl möglich sind, zeigen Maßzahlen wie SPEARMANS Rang-Korrelationskoeffizient oder PEARSONS mittlere quadratische Kontingenz.

Es soll nun die Aufgabe der vorliegenden Arbeit sein, die maßstabinvariante Korrelationstheorie systematisch und von einem einheitlichen Gesichtspunkt aus zu entwickeln. Um den Umfang unserer Abhandlung in gewissen Grenzen zu halten, werden wir uns auf zweidimensionale Verteilungen beschränken, obwohl die angewandte Methode eine Ausdehnung auf mehr als zwei Dimensionen zuläßt.

Der einfache Grundgedanke, von dem wir ausgehen werden, besteht darin, daß wir eine jede der zufälligen Variablen in der Weise transformieren, daß ihre eindimensionale Verteilung die Gleichverteilung in einem Intervall von der Länge 1 ist. Die Momente und sonstigen Parameter einer auf diese Weise „normierten“ zweidimensionalen Verteilung werden maßstabinvariant sein. Nach einer Diskussion der Gestalt einer solchen normierten Verteilung und insbesondere ihrer Summenfunktion (§ 3) behandeln wir in §§ 4—6 ihre Momente. *Die Momente kennzeichnen in ihrer Gesamtheit unabhängig vom Maßstab die zweidimensionale Verteilung.* In den meisten praktischen Anwendungen wird man natürlich danach streben, mit einem oder ganz wenigen Kennwerten auszukommen. Dem Korrelationskoeffizienten der üblichen Theorie entspricht hier der *normierte Korrelationskoeffizient* ϱ (der mit dem SPEARMANschen Rang-Korrelationskoeffizienten eng verwandt ist). Gegenüber dem gewöhnlichen Korrelationskoeffizienten r hat er die folgenden Vorzüge: er ist maßstabinvariant; er erreicht die Grenzen ± 1 nicht nur bei linearer, sondern

11) A. a. O. [10]), S. 30—31.

bei beliebiger umkehrbar eindeutiger und stetiger funktioneller Abhängigkeit der Variablen (§ 5). Ebensowenig wie beim gewöhnlichen Korrelationskoeffizienten kann mit Hilfe von ϱ festgestellt werden, ob die beiden Merkmale stochastisch unabhängig sind. In der Suche nach einem Abhängigkeitsmaß, d. h. einem Parameter, aus dessen Verschwinden die stochastische Unabhängigkeit der Variablen folgt und aus dessen Größe man auf den „Grad" der stochastischen Unabhängigkeit schließen kann, finden wir in § 7 PEARSONS *mittlere quadratische Kontingenz* φ^2 wieder, die auf eine natürliche Weise, ohne „Gewichte", eingeführt wird. Dem PEARSONschen φ^2 in gewisser Hinsicht überlegen ist ein anderes *Abhängigkeitsmaß* Φ, das in § 8 definiert wird. Φ nimmt bei umkehrbar eindeutiger und stetiger funktioneller Abhängigkeit den Wert 1, bei stochastischer Unabhängigkeit den Wert 0 und in allen übrigen Fällen einen dazwischenliegenden Wert an.

In Teil III behandeln wir die Entwicklung einer normierten Verteilungs- bzw. Summenfunktion nach orthogonalen Polynomen. Diese ermöglicht uns, Beziehungen zwischen den Parametern φ und Φ einerseits und den Momenten der normierten Verteilung andererseits herzustellen (§ 12). Aus diesen Beziehungen folgt u. a., daß das PEARSONsche φ^2 nicht kleiner als das Quadrat der dem SPEARMANschen Rang-Korrelationskoeffizienten bei kontinuierlicher Verteilung entsprechenden Größe ϱ sein kann, während $\Phi^2 \geqq \frac{5}{8}\varrho^2$ ist.

In Teil IV wird die Methode der Normierung auf den Spezialfall der zweidimensionalen GAUSSschen Verteilung angewandt.

In dieser Arbeit behandeln wir nur kontinuierliche Verteilungen. Die Übertragung unserer Ergebnisse auf diskontinuierliche und Häufigkeitsverteilungen ist auch durchgeführt und wird an anderer Stelle veröffentlicht.

Formelzusammenstellung.

Gegebene Verteilung:

$$w(\xi, \eta) \qquad (\alpha \leqq \xi \leqq \beta,\ \gamma \leqq \eta \leqq \delta),$$

$$u(\xi) = \int_{\gamma}^{\delta} w(\xi, \eta)\, d\eta, \qquad v(\eta) = \int_{\alpha}^{\beta} w(\xi, \eta)\, d\xi,$$

$$W(\xi, \eta) = \int_{\gamma}^{\eta}\int_{\alpha}^{\xi} w(\xi, \eta)\, d\xi\, d\eta,$$

$$U(\xi) = \int_{\alpha}^{\xi} u(\xi)\, d\xi, \qquad V(\eta) = \int_{\gamma}^{\eta} v(\eta)\, d\eta.$$

Normierter Korrelationskoeffizient:

$$\varrho = 12 \int_{\gamma}^{\delta}\int_{\alpha}^{\beta} \left[U(\xi) - \tfrac{1}{2}\right]\left[V(\eta) - \tfrac{1}{2}\right] w(\xi, \eta)\, d\xi\, d\eta$$

$$= 12 \int_{\gamma}^{\delta}\int_{\alpha}^{\beta} [W(\xi, \eta) - U(\xi)\, V(\eta)]\, u(\xi)\, v(\eta)\, d\xi\, d\eta.$$

$\varrho = +1$ bei monoton zunehmender funktioneller Abhängigkeit,
$\varrho = -1$ bei monoton abnehmender funktioneller Abhängigkeit,
$-1 < \varrho < +1$ in allen übrigen Fällen.

Abhängigkeitsmaß

$$\Phi^2 = 90 \int_{\gamma}^{\delta}\int_{\alpha}^{\beta} [W(\xi, \eta) - U(\xi)\, V(\eta)]^2\, u(\xi)\, v(\eta)\, d\xi\, d\eta.$$

$\Phi = 0$ bei stochastischer Unabhängigkeit,
$\Phi = 1$ bei monotoner funktioneller Abhängigkeit,
$0 < \Phi < 1$ in allen übrigen Fällen.

Gewöhnlicher Korrelationskoeffizient[12])

$$r = \frac{1}{\sqrt{\mu_{20}\, \mu_{02}}} \int_{\gamma}^{\delta}\int_{\alpha}^{\beta} [W(\xi, \eta) - U(\xi)\, V(\eta)]\, d\xi\, d\eta,$$

wo μ_{20} und μ_{02} die Streuungen von ξ und η sind.
$r = \pm 1$ bei linearer funktioneller Abhängigkeit,
$-1 < r < +1$ in allen übrigen Fällen.
r ist nicht maßstabinvariant.

Pearsons mittlere quadratische Kontingenz

$$\varphi^2 = \int_{\gamma}^{\delta}\int_{\alpha}^{\beta} \frac{[w(\xi, \eta) - u(\xi)\, v(\eta)]^2}{u(\xi)\, v(\eta)}\, d\xi\, d\eta$$

ist maßstabinvariant.

$\varphi^2 = 0$ bei stochastischer Unabhängigkeit,
$0 < \varphi^2 < \infty$ in allen übrigen Fällen.

12) Zu dieser Darstellung von r vgl. § 5.

I. Normierung einer kontinuierlichen Wahrscheinlichkeitsverteilung.

§ 1. Änderung der Verteilung bei Maßstabänderung.

Die kontinuierlichen zufälligen Variablen ξ und η mögen die Wahrscheinlichkeitsdichte (Wahrscheinlichkeitsverteilung)

$$w(\xi, \eta)$$

haben, die außerhalb des Rechtecks

$$\alpha \leqq \xi \leqq \beta, \quad \gamma \leqq \eta \leqq \delta \tag{1.1}$$

identisch verschwinden soll. Dabei können α und γ auch den Wert $-\infty$, β und δ auch den Wert $+\infty$ annehmen. Es ist

$$\int_{\gamma}^{\delta}\int_{\alpha}^{\beta} w(\xi, \eta)\, d\xi\, d\eta = 1. \tag{1.2}$$

Die eindimensionale Verteilung der Variablen ξ für sich genommen ist dann durch

$$u(\xi) = \int_{\gamma}^{\delta} w(\xi, \eta)\, d\eta \tag{1.3}$$

gegeben. $u(\xi)$ verschwindet außerhalb des Intervalls $\alpha \leqq \xi \leqq \beta$.

Entsprechend lautet die Verteilung der Variablen η für sich genommen

$$v(\eta) = \int_{\alpha}^{\beta} w(\xi, \eta)\, d\xi; \tag{1.4}$$

sie ist gleich 0 außerhalb des Intervalls $\gamma \leqq \eta \leqq \delta$.

Die Summenfunktion von ξ, also die Wahrscheinlichkeit dafür, daß ξ nicht größer als ξ_0 ist, soll mit $U(\xi_0)$ bezeichnet werden, so daß

$$U(\xi_0) = \int_{\alpha}^{\xi_0} u(\xi)\, d\xi \tag{1.5}$$

ist. Analog schreiben wir für die Summenfunktion von η

$$V(\eta_0) = \int_{\gamma}^{\eta_0} v(\eta)\, d\eta. \tag{1.6}$$

Die zweidimensionale Summenfunktion der Variablen ξ, η, also die Wahrscheinlichkeit dafür, daß $\xi \leqq \xi_0$ und $\eta \leqq \eta_0$ ist, bezeichnen wir mit $W(\xi_0, \eta_0)$ und haben

$$W(\xi_0, \eta_0) = \int_{\gamma}^{\eta_0}\int_{\alpha}^{\xi_0} w(\xi, \eta)\, d\xi\, d\eta. \tag{1.7}$$

Offenbar ist

(1. 8) $$U(\xi) = W(\xi, \delta), \quad V(\eta) = W(\beta, \eta).$$

$W(\xi, \eta)$ soll in bezug auf ξ und η stetig und bis auf höchstens abzählbar unendlich viele Ausnahmewerte von ξ und η differenzierbar vorausgesetzt werden, so daß $w(\xi, \eta)$, $u(\xi)$ und $v(\eta)$ stückweise stetig sind.

In Anknüpfung an die Ausführungen in der Einleitung wollen wir als *Maßstabänderung* eine jede Merkmaltransformation der Form

(1. 9) $$\bar{\xi} = f(\xi), \quad \bar{\eta} = g(\eta)$$

bezeichnen, sofern die Funktion $f(\xi)$ im Intervall $\alpha \leqq \xi \leqq \beta$ und die Funktion $g(\eta)$ im Intervall $\gamma \leqq \eta \leqq \delta$ eigentlich monoton zunimmt. $f(\xi)$ und $g(\eta)$ seien stetig und bis auf höchstens abzählbar unendlich viele Punkte differenzierbar, so daß die Ableitungen $f'(\xi)$ und $g'(\eta)$ in den erwähnten Intervallen positiv und endlich sind. Setzt man

$$\bar{\alpha} = f(\alpha), \quad \bar{\beta} = f(\beta), \quad \bar{\gamma} = g(\gamma), \quad \bar{\delta} = g(\delta),$$

so liegen alle Wertepaare $(\bar{\xi}, \bar{\eta})$ mit nichtverschwindender Wahrscheinlichkeitsdichte innerhalb des Rechtecks

$$\bar{\alpha} \leqq \bar{\xi} \leqq \bar{\beta}, \quad \bar{\gamma} \leqq \bar{\eta} \leqq \bar{\delta}.$$

Die Verteilung der neuen Merkmale $\bar{\xi}, \bar{\eta}$ wird im allgemeinen eine andere sein als $w(\xi, \eta)$. Sie soll mit $\bar{w}(\bar{\xi}, \bar{\eta})$ bezeichnet werden. Entsprechend sollen die übrigen Funktionen der Merkmale $\bar{\xi}, \bar{\eta}$ mit $\bar{u}(\bar{\xi})$, $\bar{W}(\bar{\xi}, \bar{\eta})$ usw. bezeichnet werden.

Da nun die Maßstabänderung (1. 9) die Gesamtheit der Wertepaare (ξ, η) mit $\xi \leqq \xi_0$, $\eta \leqq \eta_0$ genau in die Gesamtheit der Punkte $(\bar{\xi}, \bar{\eta})$ mit $\bar{\xi} \leqq \bar{\xi}_0 = f(\xi_0)$, $\bar{\eta} \leqq \bar{\eta}_0 = g(\eta_0)$ überführt, so sind auch die entsprechenden Summenwahrscheinlichkeiten einander gleich, d. h.

(1. 10) $$W(\xi_0, \eta_0) = \bar{W}(\bar{\xi}_0, \bar{\eta}_0)$$

oder

$$\int_{\gamma}^{\eta_0} \int_{\alpha}^{\xi_0} w(\xi, \eta)\, d\xi\, d\eta = \int_{\bar{\gamma}}^{\bar{\eta}_0} \int_{\bar{\alpha}}^{\bar{\xi}_0} \bar{w}(\bar{\xi}, \bar{\eta})\, d\bar{\xi}\, d\bar{\eta}.$$

Differenziert man diese Gleichung nach ξ_0 und η_0 und ersetzt sodann ξ_0 und η_0 durch ξ und η, so kommt mit Rücksicht auf (1. 9)

(1. 11) $$\bar{w}(\bar{\xi}, \bar{\eta}) = \frac{w(\xi, \eta)}{f'(\xi)\, g'(\eta)}.$$

Die Wahrscheinlichkeitsdichte der neuen Merkmale ist also gleich der ursprünglichen Wahrscheinlichkeitsdichte, dividiert durch das Produkt der Ableitungen der Transformationsfunktionen (1. 9). Wir wollen sagen: die Maßstabänderung (1. 9) führt die Verteilung $w(\xi, \eta)$ in die Verteilung $\bar{w}(\bar{\xi}, \bar{\eta})$ über.

Ähnlich findet man

$$\bar{u}(\bar{\xi}) = \frac{u(\xi)}{f'(\xi)}, \tag{1.12}$$

$$\bar{v}(\bar{\eta}) = \frac{v(\eta)}{g'(\eta)}. \tag{1.13}$$

Aus den Gleichungen (1. 11), (1. 12) und (1. 13) ersieht man sofort, daß die stochastische Unabhängigkeit eine maßstabinvariante Eigenschaft ist. Denn sind ξ und η stochastisch unabhängig, so ist $w(\xi, \eta) = u(\xi)v(\eta)$, folglich $\bar{w}(\bar{\xi}, \bar{\eta}) = \bar{u}(\bar{\xi})\bar{v}(\bar{\eta})$, so daß auch $\bar{\xi}$ und $\bar{\eta}$ stochastisch unabhängig sind. Und sind ξ und η nicht unabhängig, so sind auch $\bar{\xi}$ und $\bar{\eta}$ nicht unabhängig.

§ 2. Die Normierung.

Um die Beziehungen zwischen den Merkmalen ξ und η unabhängig von ihren Maßstäben untersuchen zu können, schlagen wir folgenden Weg ein. Wir denken uns zunächst eine wohlbestimmte eindimensionale Verteilung $e(x)$ vorgegeben. Obwohl es vom rein theoretischen Gesichtspunkte auf die Form dieser Verteilung nicht ankommt, werden wir, um von vornherein etwas Bestimmtes vor Augen zu haben, festsetzen, daß $e(x)$ die Gleichverteilung im Intervall $-\frac{1}{2} \leqq x \leqq \frac{1}{2}$ ist. Es ist also

$$\begin{cases} e(x) = 1 & \text{für} \quad -\frac{1}{2} \leqq x \leqq \frac{1}{2}, \\ e(x) = 0 & \text{sonst.} \end{cases} \tag{2.1}$$

Auf die Variable ξ mit der Verteilung $u(\xi)$ soll jetzt eine solche Transformation

$$x = x(\xi) \tag{2.2}$$

ausgeübt werden, daß x die Gleichverteilung $e(x)$ hat, und zwar soll dabei die Wahrscheinlichkeit dafür, daß $\xi \leqq \xi_0$ ist, gleich der Wahrscheinlichkeit für $x \leqq x_0 = x(\xi_0)$ sein. Diese Bedingung drückt sich unter Beachtung von (2. 1) durch

$$\int_{\alpha}^{\xi_0} u(\xi)\, d\xi = \int_{-\frac{1}{2}}^{x_0} e(x)\, dx = x_0 + \frac{1}{2}$$

aus. Da diese Beziehung für alle ξ_0 im Intervall $\alpha \leqq \xi \leqq \beta$ gelten soll, so haben wir für die Transformation (2. 2) die Schreibweise

$$x = \int_{\alpha}^{\xi} u(\xi)\, d\xi - \frac{1}{2} = U(\xi) - \frac{1}{2}. \tag{2.3}$$

x ist also der um $\frac{1}{2}$ verminderte Wert der Summenfunktion von ξ.

Abb. 1 veranschaulicht diese Transformation; die schraffierten Stücke sind flächengleich und proportional der Wahrscheinlichkeit $U(\xi_0) = x_0 + \frac{1}{2}$.

Damit die Transformation (2.3) eine Maßstabänderung ist, müssen wir voraussetzen, daß die Zahlen α und β derart bestimmt werden können, daß nicht nur, wie bisher angenommen, $u(\xi)$ außerhalb des Intervalls $\alpha \leqq \xi \leqq \beta$ verschwindet, sondern daß ferner im Innern dieses Intervalls die Wahrscheinlichkeitsdichte $u(\xi)$ durchweg größer als Null ist. Diese Voraussetzung werden wir von nun an machen, also annehmen, daß

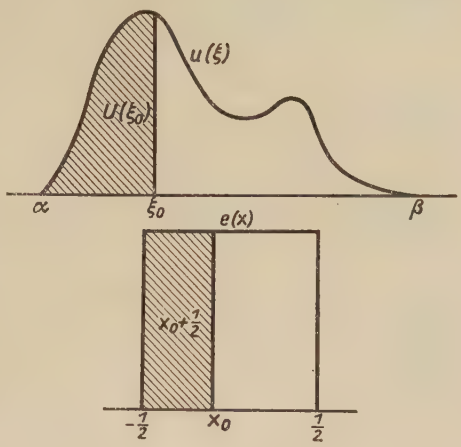

Abb. 1.

$$(2.4) \qquad \begin{cases} u(\xi) > 0 & \text{für} \quad \alpha < \xi < \beta, \\ u(\xi) = 0 & \text{für} \quad \xi < \alpha \text{ und } \xi > \beta \end{cases}$$

ist. Auch hier kann α den Wert $-\infty$, β den Wert $+\infty$ annehmen.

Wir ersehen aus dem Vorhergehenden, daß es eine und nur eine Maßstabänderung gibt, die eine Verteilung $u(\xi)$, die der Bedingung (2.4) genügt, in die Gleichverteilung $e(x)$ überführt, nämlich die durch (2.3) definierte.

Ebenso führen wir die Verteilung $v(\eta)$ durch die Transformation[13])

$$(2.5) \qquad y = \int_{\gamma}^{\eta} v(\eta)\, d\eta - \frac{1}{2} = V(\eta) - \frac{1}{2}$$

in die Gleichverteilung $e(y)$ über. Damit (2.5) eine Maßstabänderung ist, setzen wir voraus, daß

$$(2.6) \qquad \begin{cases} v(\eta) > 0 & \text{für} \quad \gamma < \eta < \delta, \\ v(\eta) = 0 & \text{für} \quad \eta < \gamma \text{ und } \eta > \delta \end{cases}$$

ist.

Die zweidimensionale Verteilung der durch die Maßstabänderungen (2.3), (2.5) eingeführten Variablen x, y sei $s(x, y)$, die Summenfunktion sei $S(x, y)$. Wir können jetzt die Gleichungen des § 1 an-

13) Die Transformation (2.3), (2.5) findet sich für den Spezialfall der Gaußschen Verteilung bei K. Pearson, a. a. O.[9]), Gleichung XII, in allgemeiner Form bei Hotelling und Pabst, a. a. O.[10]), Gleichung (5).

wenden, wobei $\bar{\xi}, \bar{\eta}$ durch $x, y, f(\xi)$ und $g(\eta)$ durch $U(\xi) - \frac{1}{2}$ und $V(\eta) - \frac{1}{2}$, $\bar{u}(\bar{\xi})$ und $\bar{v}(\bar{\eta})$ durch $e(x)$ und $e(y)$ sowie $\bar{w}$ und $\bar{W}$ durch s und S zu ersetzen sind. Dann ist nach (1. 10)

$$S(x, y) = W(\xi, \eta); \tag{2. 7}$$

im Einheitsquadrat

$$-\frac{1}{2} \leqq x \leqq \frac{1}{2}, \quad -\frac{1}{2} \leqq y \leqq \frac{1}{2} \tag{2. 8}$$

ist gemäß (1. 11), da wegen (2. 3) und (2. 5) $f'(\xi) = u(\xi)$ und $g'(\eta) = v(\eta)$ ist,

$$s(x, y) = \frac{w(\xi, \eta)}{u(\xi)\, v(\eta)}. \tag{2. 9}$$

Dabei ist

$$S(x, y) = \int_{-\frac{1}{2}}^{y} \int_{-\frac{1}{2}}^{x} s(x, y)\, dx\, dy. \tag{2. 10}$$

Zufolge (2. 1) gelten die Gleichungen

$$\begin{cases} \int_{-\frac{1}{2}}^{\frac{1}{2}} s(x, y)\, dy = 1 \quad \text{für} \quad -\frac{1}{2} \leqq x \leqq \frac{1}{2}, \\ \int_{-\frac{1}{2}}^{\frac{1}{2}} s(x, y)\, dx = 1 \quad \text{für} \quad -\frac{1}{2} \leqq y \leqq \frac{1}{2}. \end{cases} \tag{2. 11}$$

Gemäß (2. 9) ist also die Wahrscheinlichkeitsdichte der neuen Variablen x, y gleich der ursprünglichen Wahrscheinlichkeitsdichte, dividiert durch das Produkt der Wahrscheinlichkeitsdichten der einzelnen ursprünglichen Variablen.

Die rechte Seite von (2. 9) ist infolge der über $u(\xi)$ und $v(\eta)$ gemachten Voraussetzungen (2. 4) und (2. 6) im Innern des Rechtecks $\alpha \leqq \xi \leqq \beta$, $\gamma \leqq \eta \leqq \delta$ endlich, so daß $s(x, y)$ im Innern des Einheitsquadrats (2. 8) endlich ist. Auf dem Rande dieses Quadrats kann $s(x, y)$, wie wir in § 13 sehen werden, auch unendlich werden. Außerhalb des Einheitsquadrats hat man sinngemäß $s(x, y) = 0$ zu definieren, da $s(x, y)$ eine Wahrscheinlichkeitsdichte ist und das über das Einheitsquadrat erstreckte Doppelintegral von $s(x, y)$ schon den Wert 1 hat.

Nun seien $\bar{\xi}, \bar{\eta}$ zwei Variable, die aus ξ, η durch eine beliebige Maßstabänderung

$$\bar{\xi} = f(\xi), \quad \bar{\eta} = g(\eta)$$

hervorgegangen sind. Die Wahrscheinlichkeitsdichten von $\bar{\xi}$ und $\bar{\eta}$ sind durch die Gleichungen (1. 11), (1. 12), (1. 13) gegeben. Wie vorhin

die Variablen ξ, η, so sollen jetzt $\bar{\xi}, \bar{\eta}$ in die Variablen x, y mit der Gleichverteilung $e(x), e(y)$ übergeführt werden. Die zweidimensionale Verteilung der neuen Variablen sei $\bar{s}(x, y)$. Nach (2. 9) ist

$$\bar{s}(x, y) = \frac{\bar{w}(\bar{\xi}, \bar{\eta})}{\bar{u}(\bar{\xi})\,\bar{v}(\bar{\eta})}.$$

Setzt man die rechten Seiten der Gleichungen (1. 11, 12, 13) ein, so kommt

$$\bar{s}(x, y) = \frac{w(\xi, \eta)}{u(\xi)\,v(\eta)} = s(x, y).$$

Demnach werden alle zweidimensionalen Verteilungen $w(\xi, \eta)$, $\bar{w}(\bar{\xi}, \bar{\eta}), \ldots$ mit den durch (2. 4) und (2. 6) ausgedrückten Eigenschaften, die durch Maßstabänderungen auseinander hervorgehen, durch Übergang zu den Merkmalen x, y mit der eindimensionalen Gleichverteilung $e(x), e(y)$ in eine und dieselbe Verteilung übergeführt.

Wir wollen den Übergang von den Merkmalen ξ, η mit der Verteilung $w(\xi, \eta)$ zu den Merkmalen x, y mit der Verteilung $s(x, y)$ als *Normierung* der Merkmale bzw. der Verteilung bezeichnen und $s(x, y)$ die zu der Verteilung $w(\xi, \eta)$ gehörende *normierte Verteilung* nennen.

Alle Korrelations- oder Abhängigkeitsmaße oder sonstige Charakteristiken, die für die zu einer gegebenen Verteilung gehörende normierte Verteilung berechnet werden, werden für alle Verteilungen, die durch Maßstabänderung auseinander hervorgehen, die gleichen sein. *Mit Hilfe der Normierung erhält man also Charakteristiken der Verteilung, die gegenüber Maßstabänderungen invariant sind.* Die Methode der Normierung gestattet folglich, alle maßstabinvarianten Eigenschaften einer zweidimensionalen Verteilung zu erfassen.

Die Bedeutung der Methode der Normierung für die *Messung des Grades der stochastischen Abhängigkeit* tritt besonders deutlich zutage, wenn man bedenkt, daß es nur *eine* normierte Verteilung gibt, die dem Fall der *stochastischen Unabhängigkeit* der Variablen entspricht. Sind nämlich die Merkmale ξ und η stochastisch unabhängig, ist also $w(\xi, \eta) = u(\xi)v(\eta)$, so ist $s(x, y) = e(x)e(y)$ oder

$$s(x, y) = 1 \quad \text{für} \quad -\tfrac{1}{2} \leqq x \leqq \tfrac{1}{2},\ -\tfrac{1}{2} \leqq y \leqq \tfrac{1}{2},$$

$$s(x, y) = 0 \quad \text{sonst.}$$

Die einzige „unabhängige" normierte Verteilung ist also die Gleichverteilung über dem Einheitsquadrat (2. 8).

Die Summenfunktion lautet im Falle der stochastischen Unabhängigkeit

$$S_u(x, y) = \left(\tfrac{1}{2} + x\right)\left(\tfrac{1}{2} + y\right). \tag{2. 12}$$

Wie erwähnt, hätten wir für $e(x)$ statt der Gleichverteilung auch irgendeine andere eindimensionale Verteilung nehmen können; sie müßte nur eine (3. 4) entsprechende Beziehung erfüllen. Es würde z. B. naheliegen, für $e(x)$ die eindimensionale GAUSSsche Verteilung („Normalverteilung") zu nehmen.[14]) Es ist aber zu bedenken, daß wenn man für $e(x)$ keine Gleichverteilung nimmt, die mathematische Behandlung schwieriger und die Beziehungen weniger übersichtlich werden. Man wird dann die Transformation im allgemeinen nicht mehr in expliziter Form wie bei (2. 3) angeben können. An die Stelle von (2. 9) tritt

$$s(x, y) = \frac{w(\xi, \eta)}{u(\xi)\, v(\eta)}\, e(x)\, e(y).$$

Wie man sich leicht überlegt, ist die Einführung der Gleichverteilung im wesentlichen der gleiche Prozeß wie der Übergang von den ursprünglichen Werten der Variablen zu den ihrer relativen Größe entsprechenden Rangnummern.

Das Problem der Untersuchung der maßstabinvarianten Eigenschaften der zweidimensionalen Verteilungen ist durch die Transformation (2. 3, 5) auf die Behandlung von Verteilungen zurückgeführt worden, die mit Ausnahme des Einheitsquadrats (2. 8) überall in der Ebene verschwinden, wobei die eindimensionale Verteilung jedes der beiden Einzelmerkmale x und y die Gleichverteilung ist.

Durch die Voraussetzung (2. 4) haben wir den Fall ausgeschlossen, daß $w(\xi, \eta)$ in einem vertikalen Streifen durchweg verschwindet, so daß es ein Intervall $(\alpha <) \xi_1 \leqq \xi \leqq \xi_2 (< \beta)$ mit $u(\xi) = 0$ gibt. Will man die Methode der Normierung auch auf solche Fälle anwenden, so könnte man, unter der Voraussetzung, daß $u(\xi)$ sowohl in $\alpha < \xi < \xi_1$ als auch in $\xi_2 < \xi < \beta$ von 0 verschieden ist, in jedem dieser Intervalle auf ξ eine Transformation ähnlich (2. 3) ausüben. Wir werden aber diesen Ausnahmefall nicht weiter verfolgen und uns im folgenden an die Voraussetzung (2. 4, 6) halten.

Wir gehen nun dazu über, den Verlauf der Summenfunktion einer normierten Verteilung näher zu untersuchen.

---

14) Vgl. Lindeberg, a. a. O.[7]); Hotelling und Pabst, a. a. O.[10]), S. 31.

§ 3. Die normierte Summenfunktion.

Die normierte Summenfunktion

$$z = S(x, y) = \int_{-\frac{1}{2}}^{y} \int_{-\frac{1}{2}}^{x} s(x, y)\, dx\, dy \tag{3.1}$$

wird in den rechtwinkligen x, y, z-Koordinaten durch eine Fläche, die Summenfläche, dargestellt. Für uns ist nur derjenige Teil dieser Fläche von Interesse, der über dem Einheitsquadrat (2. 8) liegt.

Da $S(x, y)$ als Wahrscheinlichkeit stets zwischen 0 und 1 liegt, so verläuft die Summenfläche innerhalb des Einheitswürfels

$$-\tfrac{1}{2} \leqq x \leqq \tfrac{1}{2}, \quad -\tfrac{1}{2} \leqq y \leqq \tfrac{1}{2}, \quad 0 \leqq z \leqq 1. \tag{3.2}$$

Es läßt sich jedoch leicht einsehen, daß nur in einem bestimmten Teilstück des Einheitswürfels Werte von $S(x, y)$ liegen können. Zunächst bemerken wir, daß

$$S\left(-\tfrac{1}{2}, y\right) = 0, \quad S\left(x, -\tfrac{1}{2}\right) = 0 \tag{3.3}$$

und wegen der Gleichungen (2. 11)

$$S\left(\tfrac{1}{2}, y\right) = \tfrac{1}{2} + y, \quad S\left(x, \tfrac{1}{2}\right) = \tfrac{1}{2} + x \tag{3.4}$$

ist. Die Summenfläche schneidet also die Seitenflächen des Einheitswürfels (3. 2) stets in den gleichen Geraden, die in Abb. 2 mit a, b, c, d bezeichnet sind.

Aus (3. 4) folgt nun, da $S(x, y)$ sowohl als Funktion von x als auch als Funktion von y monoton ist,

$$\begin{cases} S(x,y) \leqq \tfrac{1}{2} + x \text{ und} \\ S(x,y) \leqq \tfrac{1}{2} + y \end{cases} \tag{3.5}$$

für alle Werte von x und y. Das bedeutet, daß die Summenfläche innerhalb der Pyramide liegt, deren Basis das Einheitsquadrat (2. 8) ist und die oben von den Ebenen $z = \frac{1}{2} + x$ und $z = \frac{1}{2} + y$ begrenzt wird (vgl. Abb. 2, wo die Schnittgerade dieser Ebenen mit e bezeichnet ist).

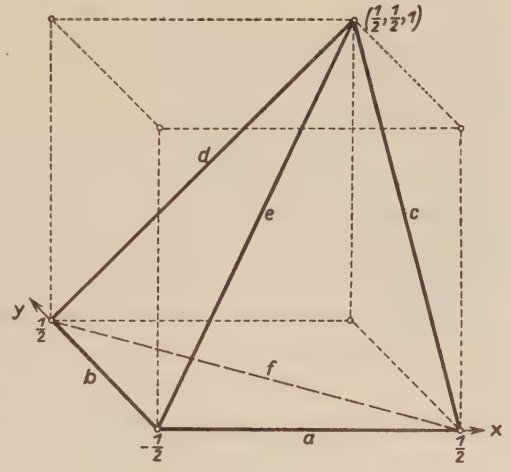

Abb. 2.

Um auch eine untere Begrenzung des Wertebereichs von $S(x, y)$ zu finden, nehmen wir zunächst eine Abschätzung einer beliebigen Summenfunktion $W(\xi, \eta)$ vor.

Die Wahrscheinlichkeit dafür, daß ein beliebiger Punkt (ξ, η) „rechts oben" vom Punkt (ξ_0, η_0) liegt, beträgt

$$\int_{\eta_0}^{\delta}\int_{\xi_0}^{\beta} w(\xi, \eta)\, d\xi\, d\eta = 1 - U(\xi_0) - V(\eta_0) + W(\xi_0, \eta_0).$$

Da diese Größe positiv ist, hat man

$$W(\xi, \eta) \geqq U(\xi) + V(\eta) - 1.$$

Speziell für eine normierte Summenfunktion gilt also, wenn man (2. 3, 5) berücksichtigt,

(3. 6) $$S(x, y) \geqq 0, \quad S(x, y) \geqq x + y.$$

Die untere Begrenzung des Wertebereichs von $S(x, y)$ wird also von den Dreiecken (a, f, b) und (c, d, f) gebildet, während die Dreiecke (a, c, e) und (b, e, d) die obere Begrenzung dieses Bereichs darstellen (vgl. Abb. 2). Die Summenfläche muß demnach innerhalb des Tetraeders mit den Kanten a, b, c, d, e, f liegen.

Diese Grenzen lassen sich nicht mehr einengen. Es gibt nämlich normierte Summenfunktionen, die der oberen Begrenzungsfläche $(a, c, e; b, e, d)$ in allen Punkten beliebig nahekommen, und andere, die in beliebiger Nachbarschaft der unteren Begrenzungsfläche $(a, f, b; c, d, f)$ verlaufen. Und zwar werden wir sehen, daß diese beiden Begrenzungsflächen die Summenflächen in gewissen Fällen der „funktionellen Abhängigkeit" der Merkmale ξ und η bilden.

Zunächst soll klargestellt werden, in welchem Sinne der Begriff der *funktionellen Abhängigkeit* hier zu verstehen ist. Wir wollen sagen, daß eine Wahrscheinlichkeitsverteilung $w(\xi, \eta)$ dem Fall der funktionellen Abhängigkeit nahekommt, wenn es eine Kurve C in der ξ, η-Ebene gibt, entlang welcher „fast" die gesamte Wahrscheinlichkeit konzentriert ist; genauer: in allen Punkten der ξ, η-Ebene, deren Entfernung von der Kurve C eine gewisse Zahl ε überschreitet, soll $w(\xi, \eta)$ sehr klein sein. Und zwar so, daß das bestimmte Integral von $w(\xi, \eta)$, erstreckt über einen Streifen längs der Kurve C, der aus denjenigen Punkten (ξ, η) besteht, deren Entfernung von C kleiner als ε ist, einen nahe an 1 liegenden Wert ergibt.[15]) Im Grenzfall der „vollkommenen" funktionellen Abhängigkeit kann man sich vorstellen,

15) Vgl. die „Korrelationsfunktion vom Streifentypus" bei K.-G. Hagström, *Bemerkungen zur Theorie der statistischen Funktionen*, Skand. Akt. Tidskr. 2 (1919), S. 204—205.

daß die gesamte nichtverschwindende Wahrscheinlichkeit sich auf der Kurve C konzentriert. Dabei muß man aber bedenken, daß in einem solchen Grenzfall die Summenfunktion $W(\xi, \eta)$ längs der Kurve C singulär und *nicht differenzierbar* sein wird, so daß auch *die Wahrscheinlichkeitsdichte nicht definiert* ist. Dagegen können die eindimensionalen Summenwahrscheinlichkeiten $U(\xi)$ und $V(\eta)$ auch in diesem Grenzfall als differenzierbar vorausgesetzt werden.

Ist η eine eindeutige und eindeutig umkehrbare und dabei stetige Funktion von ξ (also auch ξ eine ebensolche Funktion von η), so ist C eine eigentlich monotone (steigende oder fallende) Kurve.

Wenn C eine *eigentlich monoton zunehmende Kurve* ist und die Wahrscheinlichkeiten $W(\xi, \eta)$, $U(\xi)$, $V(\eta)$ in dem entsprechenden Grenzfall der funktionellen Abhängigkeit mit $W_+(\xi, \eta)$, $U_+(\xi)$, $V_+(\eta)$ bezeichnet werden, so ist, wie man aus der Bedeutung dieser Wahrscheinlichkeiten entnimmt, in allen Punkten „über" der Kurve C

$$W_+(\xi, \eta) = U_+(\xi) \tag{3. 7}$$

und $V_+(\eta) > U_+(\xi)$; „unter" der Kurve C ist

$$W_+(\xi, \eta) = V_+(\eta) \tag{3. 8}$$

und $V_+(\eta) < U_+(\xi)$. Auf der Kurve C selber gilt sowohl (3. 7) als auch (3. 8), und es ist $U_+(\xi) = V_+(\eta)$. Bezeichnet man die zugehörige normierte Summenfunktion mit $S_+(x, y)$, so ist demnach mit Rücksicht auf (2. 3, 5, 7)

$$S_+(x, y) = \begin{cases} \frac{1}{2} + x & \text{für} \quad y \geqq x, \\ \frac{1}{2} + y & \text{für} \quad y \leqq x. \end{cases} \tag{3. 9}$$

Die Summenfunktion wird also durch die Fläche $(a, c, e; b, e, d)$, die obere Begrenzungsfläche des Tetraeders in Abb. 2, dargestellt.

In dem anderen Grenzfall der umkehrbar eindeutigen stetigen funktionellen Abhängigkeit, wenn also die gesamte Wahrscheinlichkeit auf einer *eigentlich monoton abnehmenden Kurve* C' liegt, hat man, wenn die auf diesen Fall bezüglichen Funktionen mit dem Minuszeichen versehen werden, unterhalb der Kurve C'

$$W_-(\xi, \eta) = 0, \tag{3. 10}$$

wobei $V_-(\eta) < 1 - U_-(\xi)$ ist; oberhalb dieser Kurve ist

$$1 - U_-(\xi) - V_-(\eta) + W_-(\xi, \eta) = 0 \tag{3. 11}$$

und $V_-(\eta) > 1 - U_-(\xi)$. Auf C' ist sowohl (3. 10) als auch (3. 11) erfüllt, und es ist $V_-(\eta) = 1 - U_-(\xi)$. Die diesem Fall entsprechende

normierte Summenfunktion lautet folglich

$$S_-(x, y) = \begin{cases} 0 & \text{für } y \leqq -x, \\ x + y & \text{für } y \geqq -x, \end{cases} \tag{3. 12}$$

und dies entspricht der Fläche $(a, f, b; c, d, f)$.

Ist umgekehrt die normierte Summenfunktion durch (3. 9) gegeben, so daß die Summenfläche mit der oberen Tetraederbegrenzung übereinstimmt, so betrachten wir die Kurve K der ξ, η-Ebene, auf der

$$U(\xi) = V(\eta)$$

ist. Da nach Voraussetzung $u(\xi)$ im Intervall $\langle \alpha, \beta \rangle$ und $v(\eta)$ in $\langle \gamma, \delta \rangle$ von Null verschieden ist, so ist die Kurve K im Rechteck $\langle \alpha, \beta; \gamma, \delta \rangle$ eine eigentlich monoton zunehmende Kurve. Für Punkte (ξ, η) mit $V(\eta) \geqq U(\xi)$, also „oberhalb" der Kurve K, ist $W(\xi, \eta) = U(\xi)$, so daß $U(\xi_0) - W(\xi_0, \eta_0)$, die Wahrscheinlichkeit dafür, daß ein Wertepaar (ξ, η) „links oben" vom Punkt (ξ_0, η_0) liegt, für alle Punkte (ξ_0, η_0) „oberhalb" der Kurve K gleich Null ist. Entsprechend ist in allen Punkten „unterhalb" von K $V(\eta_0) - W(\xi_0, \eta_0) = 0$, also die Wahrscheinlichkeit „rechts unten" gleich Null. Dies ist gleichbedeutend damit, daß nur auf der monoton zunehmenden Kurve K eine von Null verschiedene Wahrscheinlichkeit liegen kann. η ist folglich eine umkehrbar eindeutige Funktion von ξ, und zwar nehmen ξ und η gleichzeitig zu.

Ganz analog sieht man, daß die normierte Summenfunktion (3. 12), die durch die untere Tetraederbegrenzung dargestellt wird, dem Fall der monoton abnehmenden funktionellen Abhängigkeit entspricht.

Ebenso wie es nur eine „unabhängige" normierte Verteilung gibt, gibt es also *nur je eine normierte Verteilung, die den beiden Grenzfällen der umkehrbar eindeutigen stetigen funktionellen Abhängigkeit entspricht.*

Wir haben festgestellt, daß die Summenfläche $z = S(x, y)$ innerhalb des in Abb. 2 dargestellten Tetraeders verläuft. Weitere Aufschlüsse über die Gestalt der Summenfläche erhält man, wenn man die Ableitungen von $S(x, y)$ nach x und y betrachtet. Aus (3. 1) entnimmt man zunächst, daß

$$\frac{\partial S(x, y)}{\partial x} = \int_{-\frac{1}{2}}^{y} s(x, y)\, dy$$

ist. Als Funktion von y steigt $\frac{\partial S}{\partial x}$ monoton von 0 bis 1, so daß

$$0 \leqq \frac{\partial S(x, y)}{\partial x} \leqq 1 \tag{3. 13}$$

ist. Entsprechend ist

$$0 \leqq \frac{\partial S(x, y)}{\partial y} \leqq 1. \tag{3. 14}$$

Hält man y fest, so ist $z = S(x, y)$ eine monoton steigende Kurve, deren Tangentenwinkel nirgends größer als 45^0 ist. Das Entsprechende gilt, wenn man x festhält.

Im Falle der *stochastischen Unabhängigkeit* der Merkmale ist, wie wir schon in § 2 feststellten,

$$S(x, y) = \left(\tfrac{1}{2} + x\right)\left(\tfrac{1}{2} + y\right).$$

Die dieser Funktion entsprechende Fläche, ein hyperbolisches Paraboloid, ist in Abb. 3 dargestellt.

Die Fläche

$$z = \left(\tfrac{1}{2} + x\right)\left(\tfrac{1}{2} + y\right)$$

liegt natürlich innerhalb des in Abb. 2 dargestellten Tetraeders, und zwar in dem Sinne „mitten zwischen" seinen beiden Begrenzungsflächen, daß sie das Tetraeder in zwei volumengleiche Teile zerlegt. Das sieht man deutlicher, wenn man den Wertebereich der Funktion

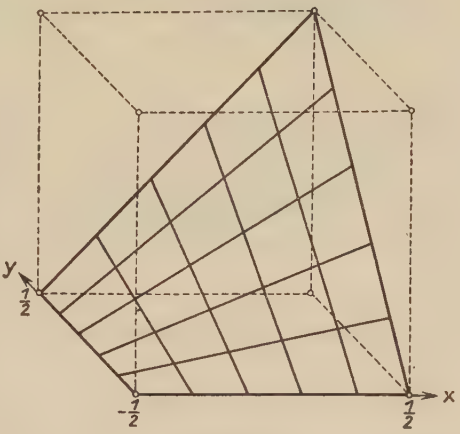

Abb. 3.

$$D(x, y) = S(x, y) - \left(\tfrac{1}{2} + x\right)\left(\tfrac{1}{2} + y\right) \tag{3. 15}$$

betrachtet, die bei stochastischer Unabhängigkeit im Einheitsquadrat (2. 8) identisch verschwindet. Aus (3. 5) und (3. 6) erhält man die Abschätzungen

$$D(x, y) \leqq \begin{cases} \left(\tfrac{1}{2} - x\right)\left(\tfrac{1}{2} + y\right) & \text{für} \quad y \leqq x, \\ \left(\tfrac{1}{2} + x\right)\left(\tfrac{1}{2} - y\right) & \text{für} \quad y \geqq x, \end{cases} \tag{3. 16}$$

$$D(x, y) \geqq \begin{cases} -\left(\tfrac{1}{2} + x\right)\left(\tfrac{1}{2} + y\right) & \text{für} \quad y \leqq -x, \\ -\left(\tfrac{1}{2} - x\right)\left(\tfrac{1}{2} - y\right) & \text{für} \quad y \geqq -x. \end{cases} \tag{3. 17}$$

Ersetzt man in (3. 17) $-x$ durch x, was eine Spiegelung bedeutet, so geht die untere Begrenzungsfläche in das Negative der oberen über. Bezeichnet man die Funktion $D(x, y)$ in den Fällen der monoton zunehmenden bzw. monoton abnehmenden funktionellen Abhängigkeit mit $D_+(x, y)$ bzw. $D_-(x, y)$, so ist folglich

$$D_-(x, y) = -D_+(-x, y). \tag{3. 18}$$

Die obere Begrenzungsfläche ist also mit der unteren kongruent (vgl. Abb. 4a und 4b). Jede dieser Flächen besteht aus zwei hyperbolischen Paraboloiden, die sich in einem scharfen Grat treffen, der bei der oberen Begrenzung über der Geraden $x = y$ liegt. Das Maximum des Grates und damit der größte Wert, den die Funktion $z = D(x, y)$ annehmen kann, ist $z = \frac{1}{4}$ und liegt bei $x = 0$, $y = 0$. Das Minimum der unteren Begrenzung beträgt $z = -\frac{1}{4}$ und liegt ebenfalls bei $x = 0$, $y = 0$. Die Ebenen $x = \text{const}$ und $y = \text{const}$ schneiden aus dem Körper, in dem die Werte von $D(x, y)$ liegen, Parallelogramme heraus.

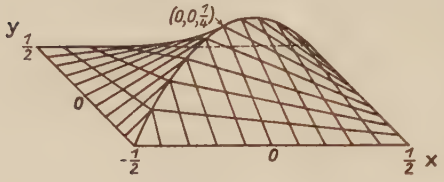

Abb. 4a.

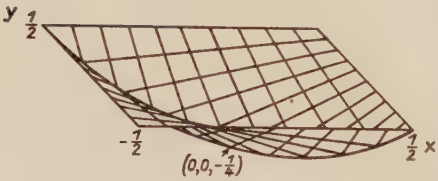

Abb. 4b.

Der Verlauf der Funktion $D(x, y)$ wird noch dadurch bestimmt, daß für die partiellen Ableitungen nach x und y die Abschätzungen

$$(3.19) \quad -\left(\frac{1}{2} + y\right) \leqq \frac{\partial D}{\partial x} \leqq \frac{1}{2} - y, \qquad -\left(\frac{1}{2} + x\right) \leqq \frac{\partial D}{\partial y} \leqq \frac{1}{2} - x$$

gelten; das folgt aus (3. 15) zusammen mit den Ungleichungen (3. 13) und (3. 14).

II. Parameter der normierten Verteilung.

§ 4. Die Momente.

Die auf den Punkt $x = 0$, $y = 0$ bezogenen *Momente der normierten Verteilung* $s(x, y)$ sind definiert durch

$$(4.1) \qquad m_{ij} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} x^i y^j s(x, y)\, dx\, dy \qquad (i = 0, 1, \ldots, j = 0, 1, \ldots).$$

Da zufolge (2. 11) $m_{10} = m_{01} = 0$ ist, so sind die m_{ij} zugleich die auf den Mittelpunkt der Verteilung bezogenen Momente.

Ferner definieren wir als *Momente der normierten Summenfunktion* $S(x, y)$ die Größen

$$(4.2) \qquad M_{ij} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} x^i y^j S(x, y)\, dx\, dy \qquad (i = 0, 1, \ldots, j = 0, 1, \ldots).$$

Geht man zu den ursprünglichen Variablen ξ, η über, so erhält man mit Rücksicht auf (2. 3, 5, 7, 9)

$$m_{ij} = \int_{\gamma}^{\delta}\int_{\alpha}^{\beta} \left[U(\xi) - \tfrac{1}{2}\right]^i \left[V(\eta) - \tfrac{1}{2}\right]^j w(\xi, \eta)\, d\xi\, d\eta,$$

$$M_{ij} = \int_{\gamma}^{\delta}\int_{\alpha}^{\beta} \left[U(\xi) - \tfrac{1}{2}\right]^i \left[V(\eta) - \tfrac{1}{2}\right]^j W(\xi, \eta)\, u(\xi)\, v(\eta)\, d\xi\, d\eta.$$

Zwischen den m_{ij} und den M_{ij} besteht die folgende Beziehung, die man durch partielle Integration von (4. 1) unter Beachtung von (2. 11) und (3. 3, 4) findet:

$$(4.3) \qquad m_{ij} = ij M_{i-1,j-1} - \frac{1}{2^{i+j}}\left(1 - \frac{1+(-1)^i}{2(i+1)} - \frac{1+(-1)^j}{2(j+1)}\right).$$

Das mittlere Glied in der Klammer nimmt den Wert 0 oder $-\frac{1}{i+1}$ an, je nachdem ob i ungerade oder gerade ist, und das Entsprechende gilt für das letzte Glied. Man hat speziell

$$m_{11} = M_{00} - \frac{1}{4}, \qquad m_{21} = 2M_{10} - \frac{1}{12},$$

$$m_{31} = 3M_{20} - \frac{1}{16}, \qquad m_{22} = 4M_{11} - \frac{1}{48}.$$

Sowohl die M_{ij} als auch die m_{ij} besitzen eine obere und eine untere Schranke. Für M_{ij} ersieht man das aus der Definitionsgleichung (4. 2), da $S(x, y)$ beschränkt ist, und wegen der Beziehung (4. 3) folgt daraus die Beschränktheit der m_{ij}.

Die Momente m_{i0} und m_{0j} sind feste, von der jeweiligen Verteilung unabhängige Zahlen; denn sie beziehen sich auf die eindimensionale Gleichverteilung zwischen $-\frac{1}{2}$ und $\frac{1}{2}$. Und zwar erhält man, da in (4. 3) im Falle $i = 0$ oder $j = 0$ an Stelle des ersten Gliedes Null zu stehen kommt,

$$(4.4) \qquad m_{i0} = m_{0i} = \begin{cases} \dfrac{1}{2^i(i+1)}, & \text{wenn } i \text{ gerade,} \\ 0, & \text{wenn } i \text{ ungerade.} \end{cases}$$

Insbesondere lauten die Mittelwertkomponenten

$$m_{10} = m_{01} = 0$$

und die Streuungskomponenten

$$m_{20} = m_{02} = \frac{1}{12}.$$

§ 5. Der normierte Korrelationskoeffizient.

Das erste von der Verteilung abhängige Moment ist m_{11}. Gleichung (4. 3) liefert die Beziehung

$$m_{11} = M_{00} - \frac{1}{4}. \tag{5. 1}$$

M_{00} ist das Volumen des Körpers, der von der Summenfläche $z = S(x, y)$ und den unteren Seitenflächen des Einheitswürfels (3. 2) eingeschlossen wird. Wir wollen diesen Körper kurz *Summenkörper* nennen. Aus dem in § 3 über die Gestalt der Summenfläche Gesagten entnimmt man sofort eine Abschätzung von M_{00}. M_{00} kann nämlich nicht größer sein als das Volumen der Pyramide, die das Einheitsquadrat (2. 8) als Basis und den Punkt $\left(\frac{1}{2}, \frac{1}{2}, 1\right)$ als Spitze hat (vgl. Abb. 2). Es ist also

$$M_{00} \leqq \frac{1}{3}.$$

Ferner ist M_{00} nicht kleiner als das Volumen der Pyramide mit der gleichen Spitze $\left(\frac{1}{2}, \frac{1}{2}, 1\right)$ und dem von den Punkten $\left(-\frac{1}{2}, \frac{1}{2}\right)$, $\left(\frac{1}{2}, -\frac{1}{2}\right)$ und $\left(\frac{1}{2}, \frac{1}{2}\right)$ der x, y-Ebene gebildeten Dreieck als Grundfläche, also

$$M_{00} \geqq \frac{1}{6}.$$

Diese beiden Grenzfälle entsprechen der monoton zunehmenden bzw. monoton abnehmenden funktionellen Abhängigkeit. Im Falle der stochastischen Unabhängigkeit ist $S(x, y) = \left(\frac{1}{2} + x\right)\left(\frac{1}{2} + y\right)$, also $M_{00} = \frac{1}{4}$. Die Größe

$$\varrho = 12\left(M_{00} - \frac{1}{4}\right) = 12\, m_{11} \tag{5. 2}$$

verschwindet also im Falle der stochastischen Unabhängigkeit und liegt zwischen den Grenzen -1 und 1. Die Werte 1 bzw. -1 nimmt ϱ dann und nur dann an, wenn zwischen den Variablen monoton zunehmende bzw. monoton abnehmende funktionelle Abhängigkeit besteht. Aus dem Verschwinden von ϱ folgt, daß der Summenkörper der betreffenden Verteilung das gleiche Volumen hat wie der Summenkörper der „unabhängigen" Verteilung.

Man kann ϱ auch in der Form

$$\varrho = 12 \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} D(x, y)\, dx\, dy = 12 \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[S(x, y) - \left(\tfrac{1}{2} + x\right)\left(\tfrac{1}{2} + y\right)\right] dx\, dy \tag{5. 3}$$

schreiben. Über den Verlauf der Funktion $D(x, y)$ war am Ende von § 3 die Rede. Aus einem positiven Wert von ϱ ist zu entnehmen, daß

das Gesamtvolumen derjenigen Körperstücke, die zwischen den positiven Teilen dieser Fläche und der x, y-Ebene liegen, größer ist als das Volumen der entsprechenden unter der x, y-Ebene liegenden Körperstücke.

Aus der Beziehung

$$\frac{m_{11}}{\sqrt{m_{20}\, m_{02}}} = 12\, m_{11} = \varrho \tag{5.4}$$

ersieht man, daß ϱ mit dem Korrelationskoeffizienten in den normierten Koordinaten identisch ist. ϱ soll deshalb kurz als *normierter Korrelationskoeffizient* bezeichnet werden. Dieser ist also gleich dem Zwölffachen des über das Einheitsquadrat erstreckten Integrals der Funktion $D(x, y)$ bzw. gleich dem um $\frac{1}{4}$ verminderten und mit 12 multiplizierten Volumen des Summenkörpers.

In den ursprünglichen Koordinaten ξ, η kann man ϱ auf folgende zwei Arten darstellen:

$$\left\{\begin{aligned} \varrho &= 12 \int_{\gamma}^{\delta}\int_{\alpha}^{\beta} \left[U(\xi) - \tfrac{1}{2}\right]\left[V(\eta) - \tfrac{1}{2}\right] w(\xi, \eta)\, d\xi\, d\eta \\ &= 12 \int_{\gamma}^{\delta}\int_{\alpha}^{\beta} [W(\xi, \eta) - U(\xi)\, V(\eta)]\, u(\xi)\, v(\eta)\, d\xi\, d\eta . \end{aligned}\right. \tag{5.5}$$

Der Unterschied von ϱ gegenüber dem gewöhnlichen Korrelationskoeffizienten r (bezogen auf die Variablen ξ, η) tritt deutlich zutage, wenn man die letztere Form von ϱ mit der folgenden Schreibweise von r vergleicht:

$$r = \frac{1}{\sqrt{\mu_{20}\, \mu_{02}}} \int_{\gamma}^{\delta}\int_{\alpha}^{\beta} [W(\xi, \eta) - U(\xi)\, V(\eta)]\, d\xi\, d\eta , \tag{5.6}$$

wo μ_{20} und μ_{02} die Streuungen von ξ und η sind. Daß diese Darstellung (die der Form (5. 3) von ϱ entspricht) mit der üblichen Definition von r übereinstimmt, verifiziert man durch partielle Integration.

Der normierte Korrelationskoeffizient unterscheidet sich also vom gewöhnlichen rein formal dadurch, daß erstens im Integranden das Produkt der Verteilungsfunktionen der Einzelvariablen als Faktor hinzutritt und zweitens vor dem Integralzeichen das reziproke geometrische Mittel der Streuungen von ξ und η durch eine feste Zahl ersetzt ist.

Diese äußeren Unterschiede haben nun, wie man aus dem Vorhergehenden entnimmt, zur Folge, daß der normierte Korrelations-

koeffizient ϱ zwei wesentliche Eigenschaften besitzt, die der gewöhnliche Korrelationskoeffizient r nicht aufzuweisen hat:

1. ϱ *ist maßstabinvariant,*

2. ϱ *erreicht die Grenzen* ± 1 *nicht nur bei linearer, sondern bei beliebiger umkehrbar eindeutiger und stetiger funktioneller Abhängigkeit.*

Der normierte Korrelationskoeffizient ϱ dürfte deshalb dem gewöhnlichen Korrelationskoeffizienten r immer dann vorzuziehen sein, wenn die Regression nicht linear oder die Gestalt der Regressionslinien unbekannt ist, man aber annehmen darf, daß diese monoton verlaufen.

ϱ ist die dem von SPEARMAN eingeführten *Rang-Korrelationskoeffizienten* bei kontinuierlichen Verteilungen entsprechende Größe.[16]).

§ 6. Die normierten Momente bei stochastischer Unabhängigkeit und bei funktioneller Abhängigkeit.

Die Werte der Momente M_{ij} der Summenfunktion bei stochastischer Unabhängigkeit und bei funktioneller Abhängigkeit ergeben sich, wenn man in (4. 2) für $S(x, y)$ die in (2. 12), (3. 9) und (3. 12) gegebenen Funktionen einsetzt. Sie sind in Tabelle 1 zusammengestellt, wobei die Indizes (u), $+$ und $-$ sich entsprechend auf die Fälle der stochastischen Unabhängigkeit, der monoton zunehmenden und der monoton abnehmenden funktionellen Abhängigkeit beziehen.

Tabelle 1.

	M_{ij}^{-}	$M_{ij}^{(u)}$	M_{ij}^{+}
i gerade, j gerade	$\frac{1}{2^{i+j+2}(i+1)(j+1)}\,\frac{i+j+2}{i+j+3}$	$\frac{1}{2^{i+j+2}(i+1)(j+1)}$	$\frac{1}{2^{i+j+2}(i+1)(j+1)}\,\frac{i+j+4}{i+j+3}$
i gerade, j ungerade	$\frac{1}{2^{i+j+2}(i+1)(j+2)}$	$\frac{1}{2^{i+j+2}(i+1)(j+2)}$	$\frac{1}{2^{i+j+2}(i+1)(j+2)}$
i ungerade, j gerade	$\frac{1}{2^{i+j+2}(i+2)(j+1)}$	$\frac{1}{2^{i+j+2}(i+2)(j+1)}$	$\frac{1}{2^{i+j+2}(i+2)(j+1)}$
i ungerade, j ungerade	$\frac{1}{2^{i+j+2}(i+2)(j+2)}\,\frac{i+j+4}{i+j+3}$	$\frac{1}{2^{i+j+2}(i+2)(j+2)}$	$\frac{1}{2^{i+j+2}(i+2)(j+2)}\,\frac{i+j+4}{i+j+3}$

Die Werte derjenigen M_{ij}, bei denen der eine Index gerade und der andere ungerade ist, stimmen in allen drei Fällen überein. In den beiden Fällen der funktionellen Abhängigkeit stimmen die M_{ij} auch dann überein, wenn sowohl i als auch j ungerade ist, und zwar sind

16) Vgl. K. Pearson, a. a. O.[9]), § 4.

sie dann größer als bei stochastischer Unabhängigkeit. Nur die Momente M, deren beide Indizes gerade sind, sind in allen drei Fällen verschieden.

Mit Hilfe der Beziehung (4. 3) lassen sich nun auch die Werte der Momente m_{ij} der normierten Verteilung bei stochastischer Unabhängigkeit berechnen sowie die Grenzwerte, denen diese Momente zustreben, wenn die Verteilung an einen der beiden Fälle der eindeutigen funktionellen Abhängigkeit nahekommt. Diese Werte sind in Tabelle 2 zusammengestellt, wobei die Indizes eine ähnliche Bedeutung wie in Tabelle 1 haben.

Tabelle 2.

	m_{ij}^-	$m_{ij}^{(u)}$	m_{ij}^+
i gerade, j gerade	$\frac{1}{2^{i+j}(i+j+1)}$	$\frac{1}{2^{i+j}(i+1)(j+1)}$	$\frac{1}{2^{i+j}(i+j+1)}$
i gerade, j ungerade	0	0	0
i ungerade, j gerade	0	0	0
i ungerade, j ungerade	$-\frac{1}{2^{i+j}(i+j+1)}$	0	$\frac{1}{2^{i+j}(i+j+1)}$

Über die relative Größe dieser Werte läßt sich dasselbe sagen wie bei den M_{ij}, bloß daß hier an die Stelle der geraden Indizes die ungeraden treten und umgekehrt.

Man beachte, daß die m_i^+ und m_{ij}^- nur von der Summe $i+j$ der Indizes abhängen.

§ 7. Die mittlere quadratische Kontingenz φ^2.

Wir stellen uns nun die Aufgabe, Größen ausfindig zu machen, die geeignet wären, als Maße der stochastischen Abhängigkeit zu dienen. Von einem solchen Abhängigkeitsmaß wird man verlangen, daß es folgende Eigenschaften besitzt:

1. Es nimmt einen festen Wert, etwa 0, dann und nur dann an, wenn die Variablen stochastisch unabhängig sind.

2. Es nimmt einen anderen festen Wert, etwa 1, dann und nur dann an, wenn zwischen den Variablen funktionelle Abhängigkeit in einem jeweils zu definierenden Sinne besteht.

3. In allen übrigen Fällen, also wenn die Variablen weder stochastisch unabhängig noch funktionell abhängig (in dem jeweils festgelegten Sinne) sind, nimmt es einen zwischen 0 und 1 liegenden Wert an.

4. Die Werte des Maßes müssen im Definitionsintervall so verteilt sein, daß es Sinn hat, aus der Größe eines Wertes auf den „Grad der stochastischen Abhängigkeit" zu schließen.

5. Das Maß hat gegenüber Änderungen des Maßstabes der Variablen invariant zu sein.

(Andere wünschenswerte Eigenschaften eines Abhängigkeitsmaßes, die mehr mit seiner praktischen Anwendung zusammenhängen, wie leichte Berechenbarkeit oder Kleinheit des wahrscheinlichen Fehlers, lassen wir hier unberücksichtigt.)

Eigenschaft 5 ist bei allen Parametern von normierten Verteilungen von selbst erfüllt.

Um eine Größe zu erhalten, die der Forderung 1 genügt, liegt es nahe, eine Funktion von ξ, η zu wählen, die im Falle der stochastischen Unabhängigkeit, und nur in diesem Falle, in einem Bereich B identisch verschwindet, und das über B erstreckte Integral des Quadrats dieser Funktion zu nehmen. Sodann wäre zu untersuchen, ob diese Größe selbst oder eine Funktion von ihr die Eigenschaften 2, 3 und 4 besitzt.

Wir haben bereits festgestellt, daß es nur eine normierte Verteilung gibt, bei der die Merkmale stochastisch unabhängig sind, nämlich die Gleichverteilung

$$s(x, y) = 1 \quad \text{für} \quad -\tfrac{1}{2} \leqq x \leqq \tfrac{1}{2}, \quad -\tfrac{1}{2} \leqq y \leqq \tfrac{1}{2}.$$

Die Funktion $s(x, y) - 1$ ist — immer innerhalb des Einheitsquadrats — dann und nur dann identisch Null, wenn stochastische Unabhängigkeit vorliegt. Folglich wird die Größe

$$\varphi^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} [s(x, y) - 1]^2 \, dx \, dy \tag{7. 1}$$

dann und nur dann verschwinden, wenn die Merkmale stochastisch unabhängig sind.

Geht man zu den ursprünglichen Variablen ξ, η über, so erhält man mit Rücksicht auf (2. 3, 5, 9)

$$\varphi^2 = \int_{\gamma}^{\delta} \int_{\alpha}^{\beta} \frac{[w(\xi, \eta) - u(\xi)\, v(\eta)]^2}{u(\xi)\, v(\eta)} \, d\xi \, d\eta. \tag{7. 2}$$

Man sieht aus dieser Gleichung, daß φ^2 gleich der von PEARSON eingeführten *mittleren quadratischen Kontingenz* (mean square contingency) in ihrer kontinuierlichen Form ist.

Man hat verschiedentlich beanstandet, daß φ^2 in der Form (7. 2) die „Gewichte" $\frac{1}{u(\xi)\,v(\eta)}$ enthält. Die einzige Rechtfertigung dieser Gewichte schien die Erlangung möglichst einfacher, praktisch brauchbarer Beziehungen zu sein.[17]) Durch (7. 1) wird nun die Größe φ^2 in einer von Gewichten freien Form definiert, die zugleich eine einfache anschauliche Deutung zuläßt. Man bestimmt zunächst die Abweichungen der gegebenen normierten Verteilung von der Gleichverteilung, die dem Fall der stochastischen Unabhängigkeit entspricht; dann ist φ^2 das über das Einheitsquadrat erstreckte Integral der Quadrate dieser Abweichungen.

Die Größe φ^2 kann bekanntlich jeden positiven Wert annehmen. Sie selbst ist deshalb als *Maß* der stochastischen Abhängigkeit schlecht geeignet. Erst eine Funktion von φ^2, die die positive Hälfte der reellen Zahlengeraden auf ein endliches Intervall projiziert, wird ein brauchbares Maß abgeben können. In der Wahl einer solchen Funktion liegt allerdings immer eine gewisse Willkür. Pearson hat hierfür den Kontingenzkoeffizienten

$$C = \sqrt{\frac{\varphi^2}{1+\varphi^2}}$$

vorgeschlagen; diese Wahl wird dadurch begründet, daß C im Falle der zweidimensionalen Gausschen Verteilung gleich dem absoluten Betrag des Korrelationskoeffizienten r ist.

Der Kontingenzkoeffizient C besitzt die Eigenschaften 2 und 3, wobei er die obere Grenze 1 auch bei nichtstetiger und nichtmonotoner funktioneller Abhängigkeit erreicht. Am schwierigsten ist die Frage zu beantworten, ob er auch Forderung 4 erfüllt. K. Pearson hat schon in seiner ersten Abhandlung über die mittlere quadratische Kontingenz[18]) darauf hingewiesen, daß φ^2 bzw. C bei Anwendung auf Häufigkeitsverteilungen unter Umständen allzu hohe Werte annehmen kann. Wir beabsichtigen, an anderer Stelle auf diese Frage zurückzukommen.

§ 8. Das Abhängigkeitsmaß Φ.

Im vorigen Paragraphen haben wir die mittlere quadratische Kontingenz φ^2 eingeführt, indem wir von der normierten Verteilung $s(x, y)$ ausgingen und sie mit der „unabhängigen" normierten Verteilung,

17) Siehe z. B. H. Cramér, *Remarks on Correlation*, Skand. Akt. Tidskr. 7 (1924), S. 230—231. — R. v. Mises, a. a. O. [5b]), S. 359; vgl. hier auch S. 315—316 die Bemerkungen über χ^2.

18) A. a. O. [5a]), S. 16.

der Gleichverteilung, verglichen. Statt dessen können wir aber auch die *Summenfunktion* $S(x, y)$ zum Ausgangspunkt nehmen und mit ihrer Hilfe eine ähnliche Größe bilden. Bei stochastischer Unabhängigkeit ist, wie wir in § 2 sahen,

$$S(x, y) = \left(\tfrac{1}{2} + x\right)\left(\tfrac{1}{2} + y\right) \quad \text{für} \quad -\tfrac{1}{2} \leqq x \leqq \tfrac{1}{2}, \; -\tfrac{1}{2} \leqq y \leqq \tfrac{1}{2}.$$

Die Funktion

$$D(x, y) = S(x, y) - \left(\tfrac{1}{2} + x\right)\left(\tfrac{1}{2} + y\right)$$

ist also dann und nur dann im Einheitsquadrat identisch Null, wenn die Merkmale ξ, η stochastisch unabhängig sind. Die Größe

$$(8.\,1)\quad \Phi^2 = 90 \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} D^2(x, y)\, dx\, dy = 90 \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[S(x, y) - \left(\tfrac{1}{2} + x\right)\left(\tfrac{1}{2} + y\right)\right]^2 dx\, dy$$

verschwindet folglich dann und nur dann, wenn stochastische Unabhängigkeit vorliegt. Dabei dient der Faktor 90 der Normierung (s. u.).

In den ursprünglichen Variablen ξ, η schreibt sich Φ^2 mit Rücksicht auf (2. 3), (2. 5) und (2. 7) in der Form

$$(8.\,2)\qquad \Phi^2 = 90 \int_{\gamma}^{\delta} \int_{\alpha}^{\beta} [W(\xi, \eta) - U(\xi)\, V(\eta)]^2\, u(\xi)\, v(\eta)\, d\xi\, d\eta\,.$$

Φ^2 hat in der Schreibweise (8. 1) eine bemerkenswerte Ähnlichkeit mit dem normierten Korrelationskoeffizienten ϱ in der Form (5. 3); während bei dem letzteren $D(x, y)$ der Integrand ist, steht bei Φ^2 das Quadrat dieser Funktion unter dem Integralzeichen. Ein entsprechender Unterschied besteht zwischen (8. 2) und der zweiten Gleichung (5. 5).

Ebenso wie das durch (7. 1) definierte φ^2 ist auch Φ^2 in der Form (8. 1) aus ungewogenen Abweichungsquadraten gebildet. Bei Übergang zu den ursprünglichen Variablen treten auch bei Φ^2 „Gewichte" auf. Während jedoch bei φ^2 die Gewichte um so größer sind, je kleiner die Wahrscheinlichkeit der Einzelmerkmale ist, fallen bei Φ^2 umgekehrt die Abweichungen derjenigen Merkmale am stärksten ins Gewicht, deren Wahrscheinlichkeit am größten ist.

Ferner unterscheidet sich Φ^2 darin von φ^2, daß es eine *endliche obere Grenze* besitzt. Das folgt aus der Beschränktheit der Summenfunktion $S(x, y)$.

Wir werden zeigen, daß Φ^2

a) sowohl bei monoton zunehmender wie bei monoton abnehmender stetiger funktioneller Abhängigkeit den Wert 1 annimmt;

b) in allen anderen Fällen kleiner als 1 ist.

Das bedeutet, daß die Größe Φ^2 bzw. ihre positive Quadratwurzel Φ die in § 7 formulierten Eigenschaften 2 und 3 besitzt, wobei die obere Grenze 1 dann und nur dann erreicht wird, wenn zwischen den Variablen umkehrbar eindeutige stetige funktionelle Abhängigkeit besteht.

Zunächst folgt aus (3. 18), daß Φ^2 bei monoton zunehmender und monoton abnehmender funktioneller Abhängigkeit einen und denselben Wert annimmt.

Bei monoton zunehmender funktioneller Abhängigkeit ist (vgl. (3. 16))

$$D_+^2(x, y) = \begin{cases} \left(\frac{1}{2} - x\right)^2 \left(\frac{1}{2} + y\right)^2 & \text{für} \quad y \leqq x, \\ \left(\frac{1}{2} + x\right)^2 \left(\frac{1}{2} - y\right)^2 & \text{für} \quad y \geqq x, \end{cases}$$

also $\Phi^2 = 1$, wie in a) behauptet.

Es soll nun gezeigt werden, daß allgemein

(8. 3) $$\Phi^2 \leqq 1$$

ist. Wir setzen zunächst

$$h(y) = \int_{-\frac{1}{2}}^{\frac{1}{2}} D^2(x, y)\, dx,$$

so daß

(8. 4) $$\Phi^2 = 90 \int_{-\frac{1}{2}}^{\frac{1}{2}} h(y)\, dy.$$

Bei festgehaltenem y bildet die durch die Ungleichungen (3. 16, 17) gegebene obere und untere Begrenzung der Funktion $D(x, y)$ ein Parallelogramm. Die erste Ungleichung (3. 19) besagt, daß die Kurve $z = D(x, y)$ nirgends stärker als die eine Seite des Begrenzungsparallelogramms wächst und nirgends stärker als die andere Parallelogrammseite fällt.

Zur Abschätzung der Funktion $D(x, y)$ schlagen wir folgenden Weg ein. Wir nehmen zunächst an, daß die Funktion $D(x, y)$ im Intervall $-\frac{1}{2} \leqq x \leqq \frac{1}{2}$ endlich viele diskrete Nullstellen hat, und ziehen durch jeden Schnittpunkt der Kurve $z = D(x, y)$ mit der x-Achse die Parallele zu derjenigen Parallelogrammseite, deren Richtungstangens dasselbe Vorzeichen hat wie der Richtungstangens von $D(x, y)$ an der betref-

fenden Nullstelle. Die Gesamtheit dieser Geraden wird einen Polygonzug $z = P(x)$ bilden, dessen Werte dem Betrage nach nicht kleiner als $D(x, y)$ sind.

Ist nämlich x_1 eine Nullstelle der Funktion $D(x, y)$, so gilt nach dem Mittelwertsatz

$$D(x, y) = (x - x_1)\left(\frac{\partial D(x, y)}{\partial x}\right)_{x=\bar{x}} \qquad (-\tfrac{1}{2} \leqq \bar{x} \leqq \tfrac{1}{2}).$$

Aus der ersten Ungleichung (3. 19) folgt dann

$$(8.5) \qquad (x - x_1)\left(\tfrac{1}{2} - y\right) \leqq D(x, y) \leqq -(x - x_1)\left(\tfrac{1}{2} + y\right) \text{ für } x \leqq x_1,$$

$$(8.6) \qquad -(x - x_1)\left(\tfrac{1}{2} + y\right) \leqq D(x, y) \leqq (x - x_1)\left(\tfrac{1}{2} - y\right) \text{ für } x \geqq x_1.$$

Wir nehmen nun an, daß die Funktion $D(x, y)$ im Innern des Intervalls $-\frac{1}{2} \leqq x \leqq \frac{1}{2}$ die $2n - 1$ Nullstellen

$$x_1, x_2, \ldots x_{2n-1}$$

hat, die der Größe nach geordnet sind. Ferner setzen wir

$$x_0 = -\tfrac{1}{2}, \qquad x_{2n} = \tfrac{1}{2}.$$

Weiterhin machen wir die Annahme, daß die Funktion $D(x, y)$ im Intervall $-\frac{1}{2} < x < x_1$ positiv ist.

In diesem Intervall ist dann nach (8. 6), wenn man x_1 durch $x_0 = -\frac{1}{2}$ ersetzt,

$$(8.7) \qquad D(x, y) \leqq (x - x_0)\left(\tfrac{1}{2} - y\right)$$

und nach (8. 5)

$$(8.8) \qquad D(x, y) \leqq -(x - x_1)\left(\tfrac{1}{2} + y\right).$$

Bezeichnet man den Punkt, in dem die oberen Grenzen in (8. 7) und (8. 8) übereinstimmen, mit a_0, so ist

$$a_0 = \tfrac{1}{2}(x_0 + x_1) + y(x_1 - x_0).$$

Im Intervall $-\frac{1}{2} \leqq x \leqq a_0$ liefert (8. 7) die bessere Abschätzung, im Intervall $a_0 \leqq x \leqq x_1$ dagegen (8. 8).

Ähnlich ist im nächsten Intervall $\langle x_1, x_2 \rangle$, wo $D(x, y)$ negativ ist,

$$D(x, y) \geqq -(x - x_1)\left(\tfrac{1}{2} + y\right),$$

$$D(x, y) \geqq (x - x_2)\left(\tfrac{1}{2} - y\right),$$

und beide Abschätzungen stimmen überein im Punkte

$$a_1 = \tfrac{1}{2}(x_1 + x_2) - y(x_2 - x_1).$$

Allgemein hat man, wenn man zu den absoluten Beträgen übergeht,

$$(8.9)\quad \begin{cases} |D(x,y)| \leqq |x - x_{2\nu}| \left(\frac{1}{2} - y\right) & \text{in} \quad \langle x_{2\nu},\ a_{2\nu}\rangle, \\ |D(x,y)| \leqq |x - x_{2\nu+1}| \left(\frac{1}{2} + y\right) & \text{in} \quad \langle a_{2\nu},\ x_{2\nu+1}\rangle, \\ |D(x,y)| \leqq |x - x_{2\nu+1}| \left(\frac{1}{2} + y\right) & \text{in} \quad \langle x_{2\nu+1},\ a_{2\nu+1}\rangle, \\ |D(x,y)| \leqq |x - x_{2\nu+2}| \left(\frac{1}{2} - y\right) & \text{in} \quad \langle a_{2\nu+1},\ x_{2\nu+2}\rangle \end{cases}$$

für $\nu = 0, 1, \ldots n-1$. Dabei ist

$$(8.10)\quad \begin{cases} a_{2\nu} = \frac{1}{2}(x_{2\nu} + x_{2\nu+1}) + y(x_{2\nu+1} - x_{2\nu}), \\ a_{2\nu+1} = \frac{1}{2}(x_{2\nu+1} + x_{2\nu+2}) - y(x_{2\nu+2} - x_{2\nu+1}). \end{cases}$$

Wenn man berücksichtigt, daß die oberen Grenzen in (8. 9) in den benachbarten Intervallen $\langle a_i,\ x_{i+1}\rangle$ und $\langle x_{i+1},\ a_{i+1}\rangle$ übereinstimmen, so ergeben sich für $D^2(x, y)$ folgende Abschätzungen:

$$D^2(x,y) \leqq \left(x + \tfrac{1}{2}\right)^2 \left(\tfrac{1}{2} - y\right)^2 \quad \text{in} \quad \left\langle -\tfrac{1}{2},\ a_0\right\rangle,$$

$$D^2(x,y) \leqq (x - x_{2\nu+1})^2 \left(\tfrac{1}{2} + y\right)^2 \quad \text{in} \quad \langle a_{2\nu},\ a_{2\nu+1}\rangle$$
$$(\nu = 0, 1, \ldots n-1),$$

$$D^2(x,y) \leqq (x - x_{2\nu+2})^2 \left(\tfrac{1}{2} - y\right)^2 \quad \text{in} \quad \langle a_{2\nu+1},\ a_{2\nu+2}\rangle$$
$$(\nu = 0, 1, \ldots n-2),$$

$$D^2(x,y) \leqq \left(x - \tfrac{1}{2}\right)^2 \left(\tfrac{1}{2} - y\right)^2 \quad \text{in} \quad \left\langle a_{2n-1},\ \tfrac{1}{2}\right\rangle.$$

Folglich ist

$$(8.11)\quad \begin{aligned} h(y) \leqq{}& \left(\tfrac{1}{2} - y\right)^2 \int_{-\frac{1}{2}}^{a_0} \left(\tfrac{1}{2} + x\right)^2 dx + \left(\tfrac{1}{2} + y\right)^2 \sum_{\nu=0}^{n-1} \int_{a_{2\nu}}^{a_{2\nu+1}} (x - x_{2\nu+1})^2 dx \\ &+ \left(\tfrac{1}{2} - y\right)^2 \sum_{\nu=0}^{n-2} \int_{a_{2\nu+1}}^{a_{2\nu+2}} (x - x_{2\nu+2})^2 dx + \left(\tfrac{1}{2} - y\right)^2 \int_{a_{2n-1}}^{\frac{1}{2}} \left(x - \tfrac{1}{2}\right)^2 dx \\ &= \bar{h}(y). \end{aligned}$$

Nach entsprechender Vereinfachung geht dieser Ausdruck über in

$$(8.12)\quad \bar{h}(y) = \frac{1}{3}\left(\frac{1}{4} - y^2\right)^2 \sum_{i=1}^{2n} (x_i - x_{i-1})^3.$$

Die Summe $\sum (x_i - x_{i-1})^3$ ist sicher kleiner als 1; denn es ist

$$\sum_{i=1}^{2n} (x_i - x_{i-1}) = x_{2n} - x_0 = 1,$$

und da die Differenzen $x_i - x_{i-1}$ alle positiv sind, somit jede kleiner als 1 ist, so ist $(x_i - x_{i-1})^3 \leqq x_i - x_{i-1}$ und $\sum (x_i - x_{i-1})^3 \leqq \sum (x_i - x_{i-1})$.

Also ist nach (8. 11)

$$h(y) \leqq \frac{1}{3}\left(\frac{1}{4} - y^2\right)^2. \tag{8. 13}$$

Ist die Anzahl der Nullstellen im Innern des Intervalls $-\frac{1}{2} \leqq x \leqq \frac{1}{2}$ gerade und sind diese mit $x_1, x_2, \ldots x_{2n}$ bezeichnet, wobei diesmal $x_{2n} < \frac{1}{2}$, $x_{2n+1} = \frac{1}{2}$ sein soll, so ist der Ausdruck (8. 12) nur dahin abzuändern, daß die Summation bis $i = 2n + 1$ zu erstrecken ist, und die übrigen Schlüsse bleiben dieselben.

Der Fall, daß $D(x, y)$ im ersten Intervall $\left\langle -\frac{1}{2}, x_1 \right\rangle$ negativ ist, geht in den bisher behandelten Fall über, wenn man $D(x, y)$ durch $-D(x, y)$ ersetzt.

Wir haben bisher angenommen, daß $D(x, y)$ als Funktion von x endlich viele Nullstellen hat. Sind unendlich viele diskrete Nullstellen vorhanden, so bleibt der Gang des Beweises im wesentlichen unverändert, nur daß in (8. 11), (8. 12) usw. an die Stelle der endlichen Summen unendliche treten.

Wenn $D(x, y)$ in einem ganzen Intervall der x-Achse verschwindet, so kann ein beliebiger Punkt dieses Intervalls als Nullstelle x_i genommen werden, und das angewandte Beweisverfahren bleibt auch auf diesen Fall anwendbar.

Damit ist Ungleichung (8. 13) allgemein bewiesen.

Aus (8. 13) folgt in Verbindung mit (8. 4)

$$\Phi^2 \leqq 30 \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\frac{1}{4} - y^2\right)^2 dy = 1.$$

Es bleibt noch zu zeigen, daß Φ^2 den Wert 1 *nur* bei monotoner funktioneller Abhängigkeit annimmt. Das ist gleichbedeutend damit, daß Φ^2 stets *kleiner* als 1 ist, wenn $D(x, y)$ weder gleich der oberen Grenze $D_+(x, y)$ noch gleich der unteren Grenze $D_-(x, y)$ ist.

Hat zunächst $D(x, y)$ mindestens eine Nullstelle im Innern des Einheitsquadrats, dann gibt es ein $y = y_1$, so daß in dem durch (8. 12) gegebenen Ausdruck für $\bar{h}(y_1)$ die Summe kleiner als 1, also $\bar{h}(y_1) < \frac{1}{3}\left(\frac{1}{4} - y_1^2\right)^2$ ist. Wegen der Stetigkeit von $D(x, y)$ ist in der

Umgebung von y_1 ebenfalls $\bar{h}(y) < \frac{1}{3}\left(\frac{1}{4} - y^2\right)^2$, so daß auch das über das Intervall $-\frac{1}{2} \leqq y \leqq \frac{1}{2}$ erstreckte Integral von $\bar{h}(y)$ kleiner als $\frac{1}{90}$ ist.

Hat aber $D(x, y)$ keine Nullstellen im Innern des Einheitsquadrats, und ist diese Funktion etwa positiv, wobei aber $D(x, y)$ nicht mit $D_+(x, y)$ übereinstimmt, so gibt es ein Wertepaar (x_0, y_0) im Innern des Einheitsquadrats, so daß $D(x_0, y_0) < D_+(x_0, y_0)$ ist. Dieselbe Ungleichung ist auch in einer Umgebung von (x_0, y_0) erfüllt, so daß das Integral von $D^2(x, y)$ wieder kleiner als $\frac{1}{90}$ ist.

Damit ist gezeigt, daß der Parameter Φ die in § 7 an ein Abhängigkeitsmaß gestellten Forderungen 1, 2, 3 und 5 erfüllt. Inwiefern er auch der Forderung 4 genügt, kann nur an Hand von Beispielen beurteilt werden. In der vorliegenden Arbeit, in der wir uns auf kontinuierliche Verteilungen beschränken, begnügen wir uns in dieser Hinsicht mit der Bestimmung von Φ für den Spezialfall der zweidimensionalen Gausschen Verteilung (§ 14).

Ein näheres Eingehen auf diese Frage wird erst in einer späteren Veröffentlichung möglich sein, in der die Übertragung der hier entwickelten Theorie auf diskontinuierliche und Häufigkeitsverteilungen durchgeführt wird. Dort soll auch gezeigt werden, daß Φ von gewissen den Kontingenzmaßen φ^2 und C anhaftenden Nachteilen frei ist.

Es sei hier noch auf die folgende Eigenschaft von Φ hingewiesen. Der Wert von Φ ändert sich nicht, wenn an Stelle der Summenfunktion $S(x, y)$, also der Wahrscheinlichkeit „links unten" von (x, y), eine der drei entsprechenden Funktionen genommen wird, die die Wahrscheinlichkeit „links oben", „rechts unten" oder „rechts oben" vom Punkt (x, y) ausdrücken. Denn bezeichnet man z. B. die Wahrscheinlichkeit „links oben" mit $S_1(x, y)$, so ist

$$S_1(x, y) = \int_{y}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{x} s(x, y)\, dx\, dy = \frac{1}{2} + x - S(x, y).$$

Im Falle der stochastischen Unabhängigkeit geht $S_1(x, y)$ über in

$$S_{1(u)}(x, y) = \frac{1}{2} + x - \left(\frac{1}{2} + x\right)\left(\frac{1}{2} + y\right),$$

und die Differenz dieser Funktionen ist

$$D_1(x, y) = S_1(x, y) - S_{1(u)}(x, y) = \left(\frac{1}{2} + x\right)\left(\frac{1}{2} + y\right) - S(x, y) = -D(x, y).$$

Es ist also $D_1^2(x, y) = D^2(x, y)$, und das über das Einheitsquadrat erstreckte Integral dieser Funktion gleich $\frac{1}{90} \Phi^2$.

Den gleichen Wert $-D(x, y)$ nimmt die $D(x, y)$ entsprechende Funktion an, wenn man von den Wahrscheinlichkeiten „rechts unten" ausgeht, während sie im Falle der Wahrscheinlichkeiten „rechts oben" gleich $D(x, y)$ ist.

III. Entwicklung normierter Verteilungs- und Summenfunktionen nach orthogonalen Polynomen.

§ 9. Entwicklung nach orthogonalen Polynomen.

Es wäre offenbar wenig zweckmäßig, die BRUNSsche Entwicklung auf normierte Verteilungen anzuwenden. Denn diese kommt in erster Linie für Verteilungen in Frage, die einigermaßen der GAUSSschen Verteilung ähneln, und das ist bei normierten Verteilungen nicht der Fall. Da der Definitionsbereich der normierten Funktionen, das Einheitsquadrat (2.8), ganz im Endlichen gelegen ist, liegt es nahe, für diese Funktionen eine *Entwicklung nach orthogonalen Polynomen* zu wählen. Eine Darstellung durch Polynome erscheint auch deshalb besonders geeignet, weil im Falle der stochastischen Unabhängigkeit sowohl die Verteilungs- als auch die Summenfunktion die Gestalt eines einfachen Polynoms, nämlich 1 bzw. $\left(\frac{1}{2} + x\right)\left(\frac{1}{2} + y\right)$, annimmt.

Unter den orthogonalen Polynomen sind die LEGENDREschen Kugelfunktionen für unsere Zwecke am geeignetsten, da bei ihrer Anwendung erstens die Entwicklungskoeffizienten sich in einfacher Weise durch die Momente ausdrücken und zweitens interessante Beziehungen zwischen den Größen φ^2 und Φ^2 einerseits und den Momenten m_{ij} und M_{ij} andererseits aufgestellt werden können. Da die LEGENDREschen Polynome im Intervall $\langle -1, 1\rangle$ orthogonal sind, werden wir sie so transformieren müssen, daß das Orthogonalitätsintervall $\left\langle -\frac{1}{2}, \frac{1}{2}\right\rangle$ ist.

Die aus den LEGENDREschen Polynomen $P_n(z)$ durch die Transformation $z = 2x$ hervorgehenden Polynome seien mit $Q_n(x)$ bezeichnet. Aus den bekannten Eigenschaften der LEGENDREschen Polynome folgt zunächst

$$Q_n(x) = \frac{1}{4^n n!} \frac{d^n (4x^2 - 1)^n}{dx^n}. \tag{9.1}$$

Ferner gelten die fundamentalen Beziehungen

$$(9.2) \qquad \int_{-\frac{1}{2}}^{\frac{1}{2}} Q_m(x)\, Q_n(x)\, dx = \begin{cases} 0 & \text{für } m \neq n, \\ \frac{1}{2n+1} & \text{für } m = n. \end{cases}$$

Man hat die Rekursionsformel

$$(9.3) \qquad (n+1)\, Q_{n+1}(x) = 2(2n+1)\, x\, Q_n(x) - n\, Q_{n-1}(x).$$

Die erste Ableitung von Q_n drückt sich durch die Q wie folgt aus:

$$(9.4) \qquad \tfrac{1}{2} Q_n'(x) = (2n-1)\, Q_{n-1}(x) + (2n-5)\, Q_{n-3}(x) + (2n-9)\, Q_{n-5}(x) + \cdots.$$

Die ersten Q_n lauten

$$(9.5) \qquad \begin{cases} Q_0 = 1, & Q_1 = 2x, & Q_2 = 6x^2 - \frac{1}{2}, \\ Q_3 = 20x^3 - 3x, & Q_4 = 70x^4 - 15x^2 + \frac{3}{8}. \end{cases}$$

Die Polynome von zwei Veränderlichen

$$(9.6) \qquad Q_{ij}(x, y) = Q_i(x) Q_j(y)$$

sind im Einheitsquadrat (2. 8) orthogonal. Und zwar folgt aus (9. 2)

$$(9.7) \qquad \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} Q_{ij}(x, y)\, Q_{kl}(x, y)\, dx\, dy = \begin{cases} 0 & \text{für } (i, j) \neq (k, l), \\ \frac{1}{(2i+1)(2j+1)} & \text{für } i = k,\ j = l. \end{cases}$$

Ebenso wie die Q_i bilden die Q_{ij} ein vollständiges Funktionensystem.

Eine im Einheitsquadrat (2. 8) erklärte Funktion $F(x, y)$ soll durch die Funktion

$$(9.8) \qquad K(x, y) = \sum_{i=0}^{k} \sum_{j=0}^{l} c_{ij}\, Q_{ij}(x, y),$$

also eine lineare Verbindung von endlich vielen orthogonalen Polynomen Q_{ij}, „im Mittel" approximiert werden, d. h. der Ausdruck

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} [F(x, y) - K(x, y)]^2\, dx\, dy$$

soll als Funktion der Koeffizienten c_{ij} ein Minimum werden. Man findet unter Berücksichtigung von (9. 7), daß die Koeffizienten die Gestalt

$$(9.9) \qquad c_{ij} = (2i+1)(2j+1) \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} F(x, y)\, Q_{ij}(x, y)\, dx\, dy$$

haben müssen.

Ist $F(x, y)$ im abgeschlossenen Einheitsquadrat stetig, so konvergiert die Reihe in (9. 8) für $k \to \infty$ und $l \to \infty$ gleichmäßig, und es ist

$$F(x, y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_{ij} Q_{ij}(x, y), \tag{9. 10}$$

wobei die c_{ij} durch (9. 9) gegeben sind. Man kann das aus dem Satz von WEIERSTRASS folgern, wonach eine jede in einem abgeschlossenen Bereich stetige Funktion gleichmäßig durch ein Polynom approximiert werden kann; denn ein Polynom kann man stets durch ein lineares Aggregat von endlich vielen $Q_{ij}(x, y)$ darstellen.

Wenn $\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} F^2(x, y)\, dx\, dy$ existiert, so läßt sich $F(x, y)$ durch einen Ausdruck der Form (9. 8) im Mittel beliebig genau approximieren.

§ 10. Polynomentwicklung von $D(x, y)$.

Wir wenden die Polynomentwicklung zunächst auf die Funktion

$$D(x, y) = S(x, y) - \left(\tfrac{1}{2} + x\right)\left(\tfrac{1}{2} + y\right) \tag{10. 1}$$

an. Da $D(x, y)$ im Innern und auf dem Rande des Einheitsquadrats (2. 8) stetig ist, so hat man

$$D(x, y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} A_{ij} Q_{ij}(x, y). \tag{10. 2}$$

Die Koeffizienten lauten gemäß (9. 9)

$$A_{ij} = (2i + 1)(2j + 1) \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} D(x, y)\, Q_{ij}(x, y)\, dx\, dy. \tag{10. 3}$$

Sétzt man

$$L_{\alpha\beta} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} x^{\alpha} y^{\beta} D(x, y)\, dx\, dy, \tag{10. 4}$$

so ist A_{ij} eine lineare Verbindung der L, und damit auch der Momente M der Summenfunktion; denn es ist offenbar

$$L_{\alpha\beta} = M_{\alpha\beta} - M^{(u)}_{\alpha\beta}, \tag{10. 5}$$

wo $M^{(u)}$ der Wert von M bei stochastischer Unabhängigkeit ist (vgl. Tabelle 1 in § 6).

Um die A durch die M auszudrücken, kann man auch folgendermaßen verfahren.

(10. 3) läßt sich wegen (10. 1) in der Form

$$A_{ij} = (2i+1)(2j+1)\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} S\,Q_{ij}\,dx\,dy$$
$$- (2i+1)(2j+1)\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}}\left(\tfrac{1}{2}+x\right)\left(\tfrac{1}{2}+y\right)Q_{ij}\,dx\,dy$$

schreiben. Nun ist, wie man aus (9. 5) entnimmt,

$$\tfrac{1}{2}+x = \tfrac{1}{2}Q_0(x) + \tfrac{1}{2}Q_1(x),$$

also mit Rücksicht auf (9. 6)

$$\left(\tfrac{1}{2}+x\right)\left(\tfrac{1}{2}+y\right) = \tfrac{1}{4}(Q_{00}+Q_{10}+Q_{01}+Q_{11}),$$

so daß wegen (9. 7)

$$(2i+1)(2j+1)\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}}\left(\tfrac{1}{2}+x\right)\left(\tfrac{1}{2}+y\right)Q_{ij}\,dx\,dy = \begin{cases}\tfrac{1}{4} \text{ für } (i,j) = \begin{cases}(0,0),\ (0,1),\\ (1,0),\ (1,1),\end{cases}\\ 0 \quad \text{sonst}\end{cases}$$

ist.

Daraus folgt, daß die Koeffizienten A_{ij} sich genau so aus den Momenten M zusammensetzen wie aus den Größen L, mit Ausnahme von A_{00}, A_{10}, A_{01} und A_{11}. Und zwar ist

$$\text{(10. 6)}\quad \begin{cases} A_{00} = L_{00} = M_{00} - \tfrac{1}{4}, \\ A_{10} = 6L_{10} = 6M_{10} - \tfrac{1}{4}, \\ A_{20} = 5\left(6L_{20} - \tfrac{1}{2}L_{00}\right) = 5\left(6M_{20} - \tfrac{1}{2}M_{00}\right), \\ A_{11} = 36L_{11} = 36M_{11} - \tfrac{1}{4}, \\ A_{30} = 7(20L_{30} - 3L_{10}) = 7(20M_{30} - 3M_{10}), \\ A_{21} = 15(12L_{21} - L_{01}) = 15(12M_{21} - M_{01}). \end{cases}$$

Im Falle der stochastischen Unabhängigkeit sind alle Koeffizienten A gleich Null.

Vergleicht man die erste Gleichung (10. 6) mit (5. 2), so sieht man, daß der erste Koeffizient A_{00} gleich einem Zwölftel des normierten Korrelationskoeffizienten ϱ ist:

$$\text{(10. 7)}\qquad A_{00} = \frac{\varrho}{12}.$$

Wegen der Beziehung (4. 3) zwischen den Momenten M und m lassen sich die A als lineare Verbindungen der Momente m darstellen. Man erhält

$$A_{00} = m_{11}, \qquad A_{10} = 3m_{21},$$
$$A_{20} = 5\left(2m_{31} - \tfrac{1}{2}m_{11}\right), \qquad A_{11} = 9m_{22} - \tfrac{1}{16},$$
$$A_{30} = 7\left(5m_{41} - \tfrac{3}{2}m_{21}\right), \qquad A_{21} = 15\left(2m_{32} - \tfrac{1}{2}m_{12}\right).$$

Wir bestimmen die ersten Koeffizienten der Entwicklung von $D(x, y)$ für die beiden Fälle der monotonen funktionellen Abhängigkeit. Setzt man die Werte der Tabelle 1 in § 6 in (10. 5) ein, so erhält man die in Tabelle 3 zusammengestellten Werte von L_{ij} (wobei die Zeichen — und + sich wie vorher auf die Fälle der monoton abnehmenden und monoton zunehmenden funktionellen Abhängigkeit beziehen).

Tabelle 3.

	L_{ij}^-	L_{ij}^+
i gerade, j gerade	$-\frac{1}{2^{i+j+2}(i+1)(j+1)(i+j+3)}$	$\frac{1}{2^{i+j+2}(i+1)(j+1)(i+j+3)}$
i gerade, j ungerade	0	0
i ungerade, j gerade	0	0
i ungerade, j ungerade	$-\frac{1}{2^{i+j+2}(i+2)(j+2)(i+j+3)}$	$\frac{1}{2^{i+j+2}(i+2)(j+2)(i+j+3)}$

Man sieht, daß die L^- und die L^+ dem Betrage nach einander gleich sind. Nun ist ein Koeffizient A_{ij}, dessen beide Indizes gerade sind, eine lineare homogene Funktion von Größen $L_{\alpha\beta}$, deren Indizes α, β ebenfalls beide gerade sind. Ebenso ist ein A_{ij} mit geradem i und ungeradem j eine lineare homogene Funktion von Größen $L_{\alpha\beta}$ mit geradem α und ungeradem β, usf. Folglich ist

$$(10.8) \qquad \begin{cases} A_{ij}^- = -A_{ij}^+, & \text{wenn } i \text{ und } j \text{ gerade,} \\ A_{ij}^- = A_{ij}^+ = 0, & \text{wenn } i \text{ gerade, } j \text{ ungerade oder } i \text{ ungerade, } j \text{ gerade,} \\ A_{ij}^- = A_{ij}^+, & \text{wenn } i \text{ und } j \text{ ungerade} \end{cases}$$

sind. Da ferner $L_{ij}^- = L_{ji}^-$ und $L_{ij}^+ = L_{ji}^+$ ist, hat man noch

$$A_{ij}^- = A_{ji}^- \quad \text{und} \quad A_{ij}^+ = A_{ji}^+.$$

Für die ersten A_{ij} bei monotoner funktioneller Abhängigkeit erhält man aus (10. 6)

$$A_{00} = \pm\frac{1}{12}, \quad A_{10} = 0, \quad A_{20} = \mp\frac{1}{12}, \quad A_{30} = 0, \quad A_{40} = 0,$$

$$A_{11} = \frac{1}{20}, \quad A_{21} = 0, \quad A_{31} = -\frac{1}{20},$$

$$A_{22} = \pm\frac{5}{42}.$$

Dabei bezieht sich das obere Vorzeichen auf den Fall der monoton zunehmenden, das untere auf den Fall der monoton abnehmenden funktionellen Abhängigkeit.

Ähnlich wie die Koeffizienten der Bruns schen Reihe lineare Verbindungen der Momente sind, sind auch die A_{ij} linear aus den normierten Momenten zusammengesetzt. So wie die Bruns schen Koeffizienten im Falle der Gauss schen Verteilung feste Werte annehmen, nehmen die A_{ij} bei stochastischer Unabhängigkeit alle denselben Wert 0 an und andere feste Werte bei monoton abnehmender oder zunehmender funktioneller Abhängigkeit. Während also die Bruns schen Koeffizienten Kriterien für die Abweichung der gegebenen Verteilung von der Gauss schen Verteilung sind, stellen die A_{ij} *Kriterien für die Abweichung von der stochastischen Unabhängigkeit und von der eindeutigen und stetigen funktionellen Abhängigkeit* dar.

Genauer gesagt können nur gewisse unter den Koeffizienten A_{ij} als solche Kriterien aufgefaßt werden. Denn einige unter ihnen, z. B. alle mit ungeradem $i+j$, verschwinden, wie erwähnt, sowohl bei stochastischer Unabhängigkeit als auch bei funktioneller Abhängigkeit. Die A_{ij}, bei denen beide Indizes i, j ungerade sind, nehmen in beiden Grenzfällen der monotonen funktionellen Abhängigkeit einen und denselben Wert an; sie können deshalb als Kriterien der Abweichung von der umkehrbar eindeutigen stetigen funktionellen Abhängigkeit als solcher gelten. Die A_{ij} endlich, bei denen sowohl i als auch j gerade ist, nehmen bei monoton zunehmender und monoton abnehmender funktioneller Abhängigkeit entgegengesetzte Werte an (sofern sie nicht, wie A_{40}, in diesen Fällen ebenfalls verschwinden); aus dem Vorzeichen und der Größe dieser Koeffizienten kann man auf die „Tendenz" der betreffenden Verteilung in Richtung auf den einen oder den anderen Grenzfall der funktionellen Abhängigkeit schließen.

Gehen wir in (10. 2) zu den ursprünglichen Variablen ξ, η über, so erhalten wir

$$W(\xi,\eta) - U(\xi)\,V(\eta) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} A_{ij}\,Q_i\left[U(\xi) - \tfrac{1}{2}\right] Q_j\left[V(\eta) - \tfrac{1}{2}\right].$$

Wir haben also die Summenfunktion $W(\xi, \eta)$ durch die Summenfunktionen $U(\xi)$ und $V(\eta)$ der Einzelvariablen ausgedrückt. Bricht man die Reihe nach endlich vielen Gliedern ab, so hat man eine näherungsweise Darstellung von $W(\xi, \eta)$ als ganze rationale Funktion von $U(\xi)$ und $V(\eta)$. Bei stochastischer Unabhängigkeit nimmt diese Beziehung die einfache Form $W(\xi, \eta) = U(\xi) V(\eta)$ an.

§ 11. Polynomentwicklung von $s(x, y)$.

Die normierte Wahrscheinlichkeitsverteilung $s(x, y)$ wird (vgl. § 9) durch die Gleichung

$$(11.1) \qquad s(x, y) = \sum_{i=0}^{k} \sum_{j=0}^{l} a_{ij} Q_{ij}(x, y) + R(x, y)$$

im Mittel approximiert, wobei R ein Rest ist. Für die Koeffizienten a_{ij} gilt nach (9. 9)

$$(11.2) \qquad a_{ij} = (2i + 1)(2j + 1) \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} s(x, y) Q_i(x) Q_j(y) \, dx \, dy .$$

Wie die A_{ij} sind auch die a_{ij} lineare Verbindungen der Momente m und damit auch der M.

Für die a_{i0} erhält man zufolge (2. 11)

$$a_{i0} = (2i + 1) \int_{-\frac{1}{2}}^{\frac{1}{2}} Q_i(x) \, dx ,$$

also wegen der Orthogonalitätsbeziehung (9. 2)

$$(11.2a) \qquad a_{00} = 1 ,$$

$$(11.2b) \qquad a_{i0} = a_{0i} = 0 \quad \text{für } i = 1, 2, \ldots$$

Für (11. 1) kann man also auch

$$(11.3) \qquad s(x, y) - 1 = \sum_{i=1}^{k} \sum_{j=1}^{l} a_{ij} Q_{ij}(x, y) + R(x, y)$$

schreiben. Im Falle der stochastischen Unabhängigkeit verschwinden hier alle a_{ij}.

Die ersten a_{ij} drücken sich wie folgt durch die Momente m aus:

$$(11.4) \qquad \begin{cases} a_{11} = 36 m_{11}, & a_{21} = 180 m_{21}, \\ a_{31} = 42 (20 m_{31} - 3 m_{11}), & a_{22} = 25 \left(36 m_{22} - \frac{1}{4}\right). \end{cases}$$

Wir bemerken, daß der erste Koeffizient a_{11} in (11. 3) mit dem normierten Korrelationskoeffizienten durch die Gleichung zusammenhängt

$$(11.5) \qquad a_{11} = 3\varrho .$$

Zwischen den a und den A besteht eine einfache Beziehung. Aus (9. 4) folgt nämlich zunächst in Verbindung mit (9. 6) für $i \geqq 1, j \geqq 1$

$$\tfrac{1}{4} Q_i'(x) Q_j'(y) = (2i-1)(2j-1) Q_{i-1,\,j-1} + (2i-1)(2j-5) Q_{i-1,\,j-3} + (2i-5)(2j-1) Q_{i-3,\,j-1} + \cdots$$

Wenn man diese Gleichung mit $D(x, y)$ multipliziert und integriert, so kommt mit Rücksicht auf (10. 3)

$$\text{(10. 6)} \qquad \tfrac{1}{4} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} D(x, y) Q_i'(x) Q_j'(y)\, dx\, dy = A_{i-1,\,j-1} + A_{i-1,\,j-3} + A_{i-3,\,j-1} \cdots$$

Partielle Integration nach x und y ergibt

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} D(x, y) Q_i'(x) Q_j'(y)\, dx\, dy = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} [s(x, y) - 1] Q_i(x) Q_j(y)\, dx\, dy.$$

Wegen (9. 2) und (11. 2) ist die rechte Seite dieser Gleichung gleich

$$\frac{a_{ij}}{(2i+1)(2j+1)} \qquad (i = 1, 2, \ldots,\ j = 1, 2, \ldots),$$

und (11. 6) geht über in

$$\text{(11. 7)} \quad \begin{cases} a_{ij} = 4(2i+1)(2j+1)(A_{i-1,\,j-1} + A_{i-1,\,j-3} + A_{i-3,\,j-1} + \cdots) \\ \quad = 4(2i+1)(2j+1) \sum_{x=0}^{\left[\frac{i-1}{2}\right]} \sum_{\lambda=0}^{\left[\frac{j-1}{2}\right]} A_{i-1-2x,\,j-1-2\lambda} \\ \qquad\qquad (i = 1, 2, \ldots,\ j = 1, 2, \ldots). \end{cases}$$

Damit ist a_{ij} durch gewisse A mit kleineren Indizes ausgedrückt.

Aus (11. 7) folgt noch die untenstehende leicht zu verifizierende Beziehung

$$\text{(11. 8)} \qquad A_{ij} = \tfrac{1}{4}\left(\frac{a_{i+1,\,j+1}}{(2i+3)(2j+3)} - \frac{a_{i+1,\,j-1}}{(2i+3)(2j-1)} - \frac{a_{i-1,\,j+1}}{(2i-1)(2j+3)} + \frac{a_{i-1,\,j-1}}{(2i-1)(2j-1)} \right),$$
$$(i, j) \neq (1, 1).$$

Zur Ableitung dieser Gleichung muß man zunächst die Annahme $i \geqq 2, j \geqq 2$ machen; sie gilt jedoch auch für die übrigen Werte von i und j mit Ausnahme von $i = j = 1$, wenn die Festsetzung

$$\text{(11. 9)} \qquad a_{-1,\,j} = a_{i,\,-1} = 0$$

getroffen wird. Dabei ist zu beachten, daß nach (11. 2b) auch $a_{i0} = a_{0i} = 0$ für $i \neq 0$ gilt, so daß, wenn i oder j den Wert 0 oder 1 annimmt, auf der rechten Seite von (11. 8) mindestens zwei von den vier Gliedern verschwinden.

Ferner ist nach (11. 7)

$$A_{11} = \frac{a_{22}}{100}.$$

Gleichung (11. 8) ergibt speziell

$$A_{00} = \frac{a_{11}}{36}, \qquad A_{10} = \frac{a_{21}}{60}, \qquad A_{20} = \frac{1}{12}\left(\frac{a_{31}}{7} - \frac{a_{11}}{3}\right),$$

$$A_{30} = \frac{1}{12}\left(\frac{a_{41}}{9} - \frac{a_{21}}{5}\right), \qquad A_{21} = \frac{1}{20}\left(\frac{a_{32}}{7} - \frac{a_{12}}{3}\right).$$

§ 12. Beziehung zwischen φ^2 und Φ^2 und den normierten Momenten.

Die Entwicklung der normierten Funktionen nach orthogonalen Polynomen gibt uns die Möglichkeit, Beziehungen zwischen dem Abhängigkeitsmaß Φ und der mittleren quadratischen Kontingenz φ^2 einerseits und den normierten Momenten M_{ij} und m_{ij} andererseits abzuleiten.

Wenn man beide Seiten von (10. 2) quadriert, über das Einheitsquadrat integriert und mit dem Faktor 90 versieht, so erhält man eine Darstellung des in (8. 1) definierten Quadrats des Abhängigkeitsmaßes Φ. Und zwar kommt, wenn man die Orthogonalitätseigenschaft der Q_{ij} berücksichtigt,

$$(12.\ 1)\quad \Phi^2 = 90 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{1}{(2i+1)(2j+1)} A_{ij}^2$$

$$= 90\left(A_{00}^2 + \frac{1}{3}A_{10}^2 + \frac{1}{3}A_{01}^2 + \frac{1}{5}A_{20}^2 + \frac{1}{9}A_{11}^2 + \frac{1}{5}A_{02}^2 + \cdots\right).$$

Da die A_{ij} lineare Verbindungen der Momente M und damit auch der m sind, so haben wir in (12. 1) das Abhängigkeitsmaß Φ durch die normierten Momente M_{ij} bzw. m_{ij} ausgedrückt.

Φ^2 ist nicht kleiner als jede Teilsumme der unendlichen Reihe in (12. 1). Dieser Tatbestand ist unter dem Namen BESSEL sche Ungleichung bekannt. Speziell ist $\Phi^2 \geqq 90 A_{00}^2$, also mit Rücksicht auf (10. 7)

$$(12.\ 2)\qquad \Phi^2 \geqq \frac{5}{8}\varrho^2, \qquad \Phi \geqq 0{,}7906\,|\varrho|.$$

Das zwischen 0 *und* 1 *liegende Abhängigkeitsmaß* Φ *ist stets größer als der mit* 0,79 *multiplizierte absolute Betrag des normierten Korrelationskoeffizienten* ϱ, *der ebenfalls zwischen* 0 *und* 1 *liegt.*

Als Beispiel, das zugleich eine gewisse Vorstellung von der Güte der Konvergenz der Reihe (12. 1) vermittelt, berechnen wir die ersten Teilsummen dieser Reihe für den Fall der monotonen funktionellen Abhängigkeit. Einen Unterschied zwischen monoton zunehmender und

abnehmender funktioneller Abhängigkeit brauchen wir nicht zu machen, da Φ in beiden Fällen den Wert 1 annimmt und die Koeffizienten A_{ij}, wie aus (10.8) folgt, dem Betrage nach einander gleich sind.

Benutzt man die in § 10 berechneten Werte der A_{ij}, so erhält man für die Summe der Reihenglieder in (12. 1) mit $i + j \leqq 2$

$$90\left(\frac{1}{144} + \frac{2}{5}\cdot\frac{1}{144} + \frac{1}{9}\cdot\frac{1}{400}\right) = \frac{9}{10},$$

also als Näherungswert für Φ die Zahl $\sqrt{0{,}9} = 0{,}949$ (statt 1). Die Teilsumme mit $i + j \leqq 4$ ist gleich $\frac{953}{980}$, die Quadratwurzel daraus gleich 0,986.

Eine entsprechende Darstellung können wir für die mittlere quadratische Kontingenz φ^2 gewinnen unter der Voraussetzung, daß der Limes der Summe in (11. 3) für $k \to \infty$ und $l \to \infty$ existiert und

$$s(x, y) - 1 = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} a_{ij}\, Q_{ij}(x, y) \tag{12. 3}$$

ist. Durch Quadrieren und Integration erhält man

$$\varphi^2 = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \frac{1}{(2i+1)(2j+1)}\, a_{ij}^2. \tag{12. 4}$$

Das Bestehen dieser Gleichung — und damit die Existenz von φ^2 — ist schon dann gewährleistet, wenn $s(x, y)$ durch die Polynomentwicklung im Mittel beliebig genau approximiert werden kann. Dann drückt (12. 4) φ^2 durch die a_{ij} und damit durch die normierten Momente m_{ij} aus.

Wenn umgekehrt φ^2 existiert (d. h. endlich ist), dann kann man $s(x, y)$ durch die Polynomentwicklung (11. 1) im Mittel beliebig genau approximieren.

Aus (12. 4) folgen wieder, wie bei Φ^2, Besselsche Ungleichungen zwischen φ^2 und den Momenten m_{ij}. Speziell ist wegen (11. 5)

$$\varphi^2 \geqq \varrho^2, \tag{12. 5}$$

also *die mittlere quadratische Kontingenz stets größer als der normierte Korrelationskoeffizient*. Für den Kontingenzkoeffizienten $C = \sqrt{\frac{\varphi^2}{1+\varphi^2}}$ folgt hieraus die Ungleichung

$$C^2 \geqq \frac{\varrho^2}{1+\varrho^2}. \tag{12. 6}$$

Die rechte Seite ist höchstens $= \frac{1}{2}$, während C^2 dem Wert 1 beliebig nahe kommen kann.

Liegt eine GAUSSsche Verteilung vor, so ist C gleich dem Betrag des Korrelationskoeffizienten r und $\varrho = \frac{6}{\pi} \arcsin \frac{r}{2}$ [19]), so daß ϱ sich nur wenig von r unterscheidet. In Ungleichung (12. 6) ist in diesem Falle der Unterschied zwischen linker und rechter Seite um so geringer, je kleiner r ist. (12. 6) besagt z. B., wenn $r^2 = 0{,}5$ ist, $0{,}50 \geqq 0{,}32$, und wenn $r^2 = 0{,}1$ ist, $0{,}100 \geqq 0{,}084$.

IV. Normierung der Gaußschen Verteilung.

§ 13. Die normierte Gaußsche Verteilung.

In diesem Paragraphen soll die Gestalt der normierten Wahrscheinlichkeitsverteilung $s(x, y)$ im Spezialfall der zweidimensionalen GAUSSschen Verteilung („Normalverteilung") näher untersucht werden.

Die Verteilung der Merkmale ξ', η' sei die GAUSSsche Verteilung

$$(13.1) \qquad w'(\xi', \eta') = \frac{1}{2\pi s t \sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left(\frac{\xi'^2}{s^2} - 2r\frac{\xi'\eta'}{st} + \frac{\eta'^2}{t^2}\right)}.$$

Hier sind s^2 und t^2 die Streuungskomponenten in bezug auf ξ' und η' und r der Korrelationskoeffizient.

Durch die Maßstabänderung

$$(13.2) \qquad \frac{\xi'}{s} = \xi, \qquad \frac{\eta'}{t} = \eta$$

geht $w'(\xi', \eta')$ zufolge (1. 11) über in die Verteilung

$$(13.3) \qquad w(\xi, \eta) = \frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}(\xi^2 - 2r\xi\eta + \eta^2)}.$$

Die Verteilungen der Einzelmerkmale ξ, η sind

$$(13.4) \qquad u(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}, \qquad v(\eta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}}.$$

Die Normierungstransformation lautet

$$(13.5) \qquad \begin{cases} x = U(\xi) - \frac{1}{2} = \frac{1}{\sqrt{2\pi}} \int_0^{\xi} e^{-\frac{\xi^2}{2}} d\xi, \\ y = V(\eta) - \frac{1}{2} = \frac{1}{\sqrt{2\pi}} \int_0^{\eta} e^{-\frac{\eta^2}{2}} d\eta. \end{cases}$$

19) K. Pearson, a. a. O.[9]), S. 12.

Die Wahrscheinlichkeitsverteilung der normierten Variablen x, y ist nach (2.9) gleich

$$s(x, y) = \frac{w[\xi(x), \eta(y)]}{u[\xi(x)]\, v[\eta(y)]}.$$

Wir betrachten zunächst die Funktion

$$s[x(\xi), y(\eta)] = \frac{w(\xi, \eta)}{u(\xi)\, v(\eta)}$$

in den Koordinaten ξ, η. Es ist

$$s(x, y) = \frac{1}{\sqrt{1-r^2}}\, e^{-\frac{r}{2(1-r^2)}(r\xi^2 - 2\xi\eta + r\eta^2)}. \tag{13.6}$$

Für $r = 0$ ist $s(x, y)$ konstant gleich 1, wie es bei stochastischer Unabhängigkeit auch sein soll.

Sei jetzt $r \neq 0$. Dann kann man (13.6) in der Form

$$s(x, y) = \frac{1}{\sqrt{1-r^2}}\, e^{-\frac{r^2}{2(1-r^2)} Q(\xi, \eta)} \tag{13.7}$$

schreiben, wo

$$Q(\xi, \eta) = \xi^2 - \frac{2}{r}\xi\eta + \eta^2 \tag{13.8}$$

ist. Die Diskriminante $1 - \frac{1}{r^2}$ der quadratischen Form Q ist negativ, da der Fall $r^2 = 1$ auszuschließen ist. Die Horizontalschnitte (Höhenlinien) $s[x(\xi), y(\eta)] = \text{const}$ oder

$$\xi^2 - \frac{2}{r}\xi\eta + \eta^2 = C \tag{13.9}$$

bilden eine Schar von Hyperbeln. Im Falle $C = 0$ artet die Hyperbel in das Geradenpaar

$$\eta = \frac{r}{1 \pm \sqrt{1-r^2}}\xi = \frac{1 \mp \sqrt{1-r^2}}{r}\xi \tag{13.10}$$

aus.

Die Geraden (13.10) sind die gemeinsamen Asymptoten der Hyperbelschar (13.9). Ihre Richtungstangenten haben dasselbe Vorzeichen, und sie liegen symmetrisch zur Geraden $\eta = \xi$.

Auf den Asymptoten (13.10) ist $s(x, y)$ konstant gleich $\frac{1}{\sqrt{1-r^2}}$. Zwischen ihnen, also für die Werte von η, die zwischen

$$\frac{r}{1+\sqrt{1-r^2}}\xi \quad \text{und} \quad \frac{r}{1-\sqrt{1-r^2}}\xi$$

liegen, ist $Q(\xi, \eta) < 0$, also $s(x, y) > \frac{1}{\sqrt{1-r^2}}$. „Außerhalb" der Asymptoten ist $s(x, y) < \frac{1}{\sqrt{1-r^2}}$.

Auf einer Geraden durch den Ursprung

(13. 11) $$\eta = k\xi$$

ist $$Q(\xi, \eta) = \xi^2\left(1 - \frac{2}{r}k + k^2\right).$$

Der Ausdruck in der Klammer verschwindet, wenn k einen der Werte

$$k_1 = \frac{r}{1+\sqrt{1-r^2}}, \qquad k_2 = \frac{r}{1-\sqrt{1-r^2}}$$

annimmt, also die Gerade (13. 11) mit einer der Asymptoten (13. 10) zusammenfällt. Wenn k zwischen k_1 und k_2 liegt, also die Gerade (13. 11) zwischen den Asymptoten verläuft, so ist der Ausdruck $1 - \frac{2}{r}k + k^2$ positiv, und wenn k außerhalb dieser Werte liegt, ist er negativ. Geht man also auf einer Geraden (13. 11) nach $\pm\infty$, so wächst $s(x, y)$ über alle Grenzen, wenn k zwischen k_1 und k_2 liegt; liegt aber k außerhalb dieses Intervalls, so strebt $s(x, y)$ gegen Null.

$s(x, y)$ nähert sich auch dann der Null, wenn man auf einer Geraden $\xi = \text{const}$ oder $\eta = \text{const}$ nach $\pm\infty$ geht. Daraus geht hervor, daß $s(x, y)$ bei festgehaltenem $\eta = \eta_0$ und ebenso bei festgehaltenem $\xi = \xi_0$ ein Maximum haben muß. Diese Extrema stimmen mit denjenigen von $Q(\xi, \eta_0)$ bzw. $Q(\xi_0, \eta)$ überein. Und zwar findet man, daß das Maximum von $s[x(\xi), y(\eta)]$ bei festem $\eta = \eta_0$ an der Stelle

$$\xi = \frac{1}{r}\eta_0$$

und bei festem $\xi = \xi_0$ an der Stelle

$$\eta = \frac{1}{r}\xi_0$$

liegt. Die Maxima befinden sich also gerade auf den Regressionslinien.

Die Regressionslinien haben ja auch bei der Verteilungsfunktion $w(\xi, \eta)$ dieselbe Bedeutung als Ort der Maxima. Der Unterschied besteht bloß darin, daß das Maximum von $w(\xi, \eta)$ bei festgehaltenem $\eta = \eta_0$ an der Stelle

$$\xi = r\eta_0$$

und bei festem $\xi = \xi_0$ im Punkte

$$\eta = r\xi_0$$

liegt. Die Rolle der beiden Regressionsgeraden

(13. 12) $$\eta = r\xi \quad \text{und} \quad \eta = \frac{1}{r}\xi$$

ist also in den beiden Fällen vertauscht.

Die Richtungstangenten der beiden Regressionsgeraden (13. 12) und der beiden Asymptoten (13. 10) haben alle vier dasselbe Vorzeichen. Für sie gelten die Ungleichungen

$$\left|\frac{r}{1+\sqrt{1-r^2}}\right| < |r| < \left|\frac{1}{r}\right| < \left|\frac{1+\sqrt{1-r^2}}{r}\right|.$$

Man sieht daraus, daß die Regressionsgeraden „zwischen" den Asymptoten liegen.

Wir gehen nun zu den normierten Variablen x, y über.

Durch die Transformation (13. 5) wird die gesamte ξ, η-Ebene auf das Einheitsquadrat $-\frac{1}{2} \leqq x \leqq \frac{1}{2}$, $-\frac{1}{2} \leqq y \leqq \frac{1}{2}$ abgebildet. Es soll die Gestalt derjenigen Kurven der x, y-Ebene untersucht werden, in die die durch den Ursprung der ξ, η-Ebene verlaufenden Geraden übergehen.

Aus (13. 5) sieht man zunächst, daß die Gerade $\xi = 0$ in $x = 0$ übergeht, ebenso die Gerade $\eta = 0$ in $y = 0$.

Die Geraden

$$\eta = k\xi$$

gehen bei dieser Abbildung in die Kurven

$$y = f(x) = \frac{1}{\sqrt{2\pi}} \int_0^{k\xi(x)} e^{-\frac{\eta^2}{2}} d\eta \tag{13. 13}$$

über, wobei $\xi(x)$ durch die erste Gleichung (13. 5) gegeben ist.

Die Kurven $y = f(x)$ haben alle den Punkt $(x, y) = (0, 0)$ gemeinsam. Bei positivem k geht $y = f(x)$ durch die Eckpunkte $\left(-\frac{1}{2}, -\frac{1}{2}\right)$ und $\left(\frac{1}{2}, \frac{1}{2}\right)$ des Einheitsquadrats und steigt zwischen diesen Punkten monoton. Bei negativem k beginnt die Kurve im Eckpunkt $\left(-\frac{1}{2}, \frac{1}{2}\right)$, fällt monoton und endet im Punkte $\left(\frac{1}{2}, -\frac{1}{2}\right)$.

Um die Gestalt dieser Kurven deutlicher zu erkennen, bilden wir die ersten beiden Ableitungen von $f(x)$ nach x:

$$f'(x) = ke^{-\frac{1}{2}(k^2-1)\xi^2}, \tag{13. 14}$$

$$f''(x) = \sqrt{2\pi}k(1-k^2)\xi e^{-\frac{1}{2}(k^2-2)\xi^2}. \tag{13. 15}$$

Aus (13. 15) ersieht man zunächst, daß, von dem bereits erledigten Fall $k = 0$ abgesehen, die Kurve $y = f(x)$ dann und nur dann eine Gerade ist, wenn $k = \pm 1$ ist, die Gleichungen der Geraden in der ξ, η-Ebene also

$$\eta = \pm \xi$$

lauten.

Als Abbildungen dieser Geraden in den x, y-Koordinaten erhält man, da $f'(x) = \pm 1$ und $f(0) = 0$ ist, das Geradenpaar

$$y = \pm x.$$

Ist $k \neq \pm 1$ und $\neq 0$, so ersieht man aus (13. 15), daß die Kurve $y = f(x)$ an der Stelle $\xi = 0$ oder $x = 0$ einen Wendepunkt hat. Im übrigen hängt der Verlauf der Kurve davon ab, ob k positiv oder negativ und ob $|k|$ größer oder kleiner als 1 ist. Die Vorzeichen der Ableitungen f' und f'' in diesen vier Fällen sowie die Grenzwerte

$$f'\left(\tfrac{1}{2}\right) = \lim_{x \to \frac{1}{2}} f'(x) = \lim_{x \to -\frac{1}{2}} f'(x) = f'\left(-\tfrac{1}{2}\right)$$

sind in Tabelle 4 zusammengestellt. Auch die Werte $f\left(-\frac{1}{2}\right)$ und $f\left(\frac{1}{2}\right)$ sind noch einmal angegeben.

Tabelle 4.

	$f\left(-\frac{1}{2}\right)$	$f\left(\frac{1}{2}\right)$	$f'\left(\frac{1}{2}\right) = f'\left(-\frac{1}{2}\right)$	Vorzeichen von $f'(x)$	Vorzeichen von $f''(x)$ $x < 0$	$x > 0$
$1 < k$	$-\frac{1}{2}$	$\frac{1}{2}$	0	+	+	−
$0 < k < 1$	$-\frac{1}{2}$	$\frac{1}{2}$	$+\infty$	+	−	+
$-1 < k < 0$	$\frac{1}{2}$	$-\frac{1}{2}$	$-\infty$	−	+	−
$k < -1$	$\frac{1}{2}$	$-\frac{1}{2}$	0	−	−	+

Im Fall $k > 1$ z. B. beginnt die Kurve $f(x)$ im Punkt $(x, y) = \left(-\frac{1}{2}, -\frac{1}{2}\right)$ mit der Steigung 0, der Richtungstangens wächst dann monoton, bis er im Punkte (0,0) den Höchstwert k erreicht; von hier ab wird die Steigung wieder geringer, bis sie in $\left(\frac{1}{2}, \frac{1}{2}\right)$ wieder 0 wird. Entsprechend ersieht man aus der Tabelle den Verlauf der Kurve in den übrigen Fällen.

In Abb. 5 sind die Bilder $y = f(x)$ der Geraden $\eta = \frac{5}{3}\xi$ und $\eta = \frac{3}{5}\xi$ punktiert eingezeichnet. Die mit der Zahl 1,250 versehenen Kurven sind die Bilder der Geraden $\eta = 3\xi$ und $\eta = \frac{1}{3}\xi$.

Läßt man bei festem ξ die Richtungskonstante k gegen $+\infty$ gehen, so folgt aus (13. 13)

$$f(x) \to -\tfrac{1}{2} \quad \text{für} \quad x < 0,$$

$$f(x) \to \tfrac{1}{2} \quad \text{für} \quad x > 0.$$

Die Geradenbilder (13. 13) nähern sich also mit wachsender Richtungskonstante k einer Treppenkurve. Ähnlich hat man, wenn k gegen $-\infty$ strebt,

$$f(x) \longrightarrow \frac{1}{2} \quad \text{für} \quad x < 0,$$

$$f(x) \longrightarrow -\frac{1}{2} \quad \text{für} \quad x > 0.$$

Die Bildkurven der Asymptoten (13. 10) und der Hyperbeln (13. 9) sind die Höhenlinien der Verteilungsfläche $z = s(x, y)$ und damit die Isostochen (Linien gleicher Wahrscheinlichkeitsdichte) dieser Verteilung.

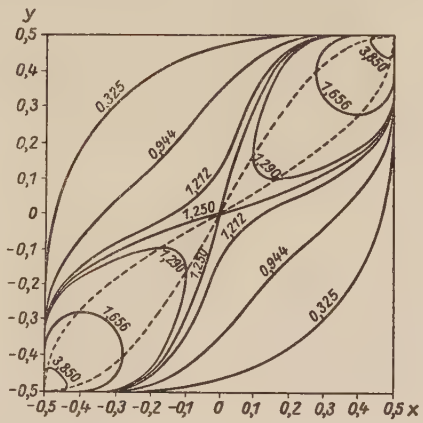

Abb. 5.

Die Richtungsfaktoren

$$\frac{r}{1+\sqrt{1-r^2}} \quad \text{und} \quad \frac{r}{1-\sqrt{1-r^2}}$$

der Asymptoten haben beide dasselbe Vorzeichen, der eine ist aber dem Betrage nach kleiner als 1, der andere größer. Demnach gehen beide Asymptotenbilder durch dieselben Ecken des Einheitsquadrats und bilden zusammen eine Figur, die einer schräggestellten 8 gleicht. Die „8" ist nach rechts oder nach links geneigt, je nachdem der Korrelationskoeffizient r positiv oder negativ ist. Die „8" ist um so schmaler, je näher $|r|$ an 1 ist, und um so breiter, je weniger r sich von 0 unterscheidet. Im Innern der „8" ist die Wahrscheinlichkeitsdichte $s(x, y)$ größer als $\frac{1}{\sqrt{1-r^2}}$, außerhalb kleiner.

Die übrigen Isostochen, die Bildkurven der Hyperbeln (13. 9), schmiegen sich mit $x \longrightarrow -\frac{1}{2}$ und $x \longrightarrow \frac{1}{2}$ den Asymptotenbildern an. Alle Isostochen münden also in den beiden Endpunkten, durch die die Asymptotenbilder gehen.

Die Bilder der Regressionsgeraden (13. 12) bilden ebenfalls eine „8", die in die „8" der Asymptotenbilder eingebettet ist. Eine Ausnahme bildet nur der Fall $r = 0$, wo sie mit den Geraden $x = 0$ und $y = 0$ übereinstimmen.

In Abb. 5 sind für den Fall $r = 0{,}6$ die Isostochen von $s(x, y)$ eingezeichnet, für die die Konstante C in (13. 9) die Werte

$$C = -4, \quad -1, \quad -\frac{1}{9}, \quad 0, \quad \frac{1}{9}, \quad 1, \quad 4$$

annimmt. Auf diesen Isostochen ist in derselben Reihenfolge

$$s(x, y) = 3{,}850, \quad 1{,}656, \quad 1{,}290, \quad 1{,}250, \quad 1{,}212, \quad 0{,}944, \quad 0{,}325.$$

Die punktierten Linien sind die Bilder der Regressionsgeraden.

Die Verteilungsfläche $z = s(x, y)$ hat bei $x = y = 0$ einen Sattelpunkt. Ist r positiv, so ist am Rande des Einheitsquadrats $s(x, y) = 0$ mit Ausnahme der Punkte $\left(-\frac{1}{2}, -\frac{1}{2}\right)$ und $\left(\frac{1}{2}, \frac{1}{2}\right)$, die singulär sind. In diesen Punkten nimmt $s(x, y)$ jeden beliebigen positiven Wert an, wenn man sich ihnen auf der entsprechenden Isostoche nähert. Ist r negativ, so hat die Verteilungsfläche ein ähnliches Aussehen, bloß mit dem Unterschied, daß an die Stelle der Punkte $\left(-\frac{1}{2}, -\frac{1}{2}\right)$ und $\left(\frac{1}{2}, \frac{1}{2}\right)$ die Singularitäten $\left(-\frac{1}{2}, \frac{1}{2}\right)$ und $\left(\frac{1}{2}, -\frac{1}{2}\right)$ treten.[20])

§ 14. Φ bei der Gaußschen Verteilung.

Eine Reihe von Parametern der normierten GAUSSschen Verteilung ist bereits von K. PEARSON in seinen Arbeiten über die mittlere quadratische Kontingenz und den SPEARMANschen Rang-Korrelationskoeffizienten berechnet worden. Insbesondere gilt bei GAUSSscher Verteilung (13. 1) für die mittlere quadratische Kontingenz[21])

$$\varphi^2 = \frac{r^2}{1 - r^2}; \tag{14. 1}$$

ferner fand K. PEARSON[22]) für den normierten Korrelationskoeffizienten

$$\varrho = \frac{6}{\pi} \arcsin \frac{r}{2}. \tag{14. 2}$$

Auch einige höhere Momente der normierten GAUSSschen Verteilung sind von PEARSON als Funktionen von r bestimmt worden.[23])

Im folgenden soll auch das Abhängigkeitsmaß Φ durch r ausgedrückt werden.

20) Bei „Student", *An Experimental Determination of the Probable Error of Dr. Spearman's Correlation Coefficients*, Biometrika 13 (1920—21), S. 278, Fußnote, findet sich ein kurzer Hinweis auf die Gestalt der „Rang-Korrelationsfläche", die aus „zwei verhältnismäßig steilen, durch einen Sattel verbundenen Graten" besteht, wobei „die Grate einen schiefen Querschnitt haben". Aus dem Text ist nicht klar ersichtlich, ob empirische „Rangverteilungen" gemeint sind oder die normierte Gaußsche Verteilung.

21) A. a. O.[5a]), S. 8.

22) A. a. O.[9]), S. 12.

23) A. a. O.[9]), S. 25—32.

Zu der Gausschen Verteilung der Merkmale ξ, η mit den Mittelwerten 0 und den Streuungen 1 gehört die Summenfunktion

$$(14.3) \qquad W(\xi,\eta) = \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{\eta}\int_{-\infty}^{\xi} e^{-\frac{1}{2(1-r^2)}(\xi^2 - 2r\xi\eta + \eta^2)}\, d\xi\, d\eta .$$

Die Summenfunktionen der Einzelvariablen sind

$$(14.4) \qquad U(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi} e^{-\frac{\xi^2}{2}} d\xi, \qquad V(\eta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\eta} e^{-\frac{\eta^2}{2}} d\eta .$$

K. Pearson hat gezeigt[24]), daß die Funktion $W - UV$ in diesem Fall sich folgendermaßen nach Potenzen von r entwickeln läßt:

$$(14.5) \qquad W(\xi,\eta) - U(\xi)\,V(\eta) = \frac{1}{2\pi} e^{-\frac{1}{2}(\xi^2+\eta^2)} \sum_{n=1}^{\infty} \frac{r^n}{n!} p_{n-1}(\xi)\, p_{n-1}(\eta) .$$

Dabei sind die p_n Polynome, die durch die Beziehungen

$$(14.6) \qquad \begin{cases} p_0(\xi) = 1, \qquad p_1(\xi) = \xi, \\ p_n(\xi) = \xi\, p_{n-1}(\xi) - (n-1)\, p_{n-2}(\xi) \end{cases}$$

definiert sind. Die p_n hängen aufs engste mit den Hermiteschen Polynomen

$$(14.7) \qquad h_n(z) = (-1)^n e^{z^2} \frac{d^n e^{-z^2}}{dz^n}$$

zusammen. Es ist nämlich

$$(14.8) \qquad p_n(\xi) = \frac{1}{(\sqrt{2})^n} h_n\left(\frac{\xi}{\sqrt{2}}\right).$$

Die Entwicklung von $W - UV$ nach Potenzen von r ist demnach gleichzeitig die Entwicklung dieser Funktion in eine Brunssche Reihe.

---

24) K. Pearson, *On the Correlation of Characters not Quantitatively Measurable*, Phil. Trans. Roy. Soc. London, Ser. A, 195 (1901), Gleichung (XIX). In Pearsons Schreibweise lautet (14. 5)

$$\frac{ad - bc}{N^2 HK} = \sum_{1}^{\infty} \frac{r^n}{n!} \bar{v}_{n-1}\, \bar{w}_{n-1} .$$

Die Pearsonschen Bezeichnungen hängen mit den von uns verwendeten wie folgt zusammen:

$$\frac{a}{N} = W, \qquad \frac{b}{N} = V - W, \qquad \frac{c}{N} = U - W, \qquad \frac{d}{N} = 1 - U - V + W,$$

$$H = \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}, \qquad K = \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}}, \qquad \bar{v}_n = p_n(\xi), \qquad w_n = p_n(\eta) .$$

Die p_n genügen der Gleichung

$$p_n'(\xi) = n p_{n-1}(\xi), \tag{14. 9}$$

die man mit Hilfe von (14. 8) aus der für HERMITE sche Polynome geltenden Beziehung

$$h_n'(z) = 2n h_{n-1}(z)$$

folgert.

Wenn man beide Seiten der Gleichung (14. 5) quadriert und mit den durch (13. 4) gegebenen Funktionen $u(\xi)$, $v(\eta)$ multipliziert, so kommt

$$u(\xi)\, v(\eta)\, [W(\xi, \eta) - U(\xi)\, V(\eta)]^2 = \frac{1}{8\pi^3} e^{-\frac{3}{2}(\xi^2+\eta^2)} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{r^{i+j}}{i!\, j!} p_{i-1}(\xi)\, p_{j-1}(\xi)\, p_{i-1}(\eta)\, p_{j-1}(\eta).$$

Das über die ganze ξ, η-Ebene erstreckte Integral der linken Seite dieser Gleichung ist nach (8. 2) gleich $\frac{1}{90} \Phi^2$; man erhält also

$$\Phi^2 = \frac{15}{2\pi^2} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{r^{i+j}}{i!\, j!} k_{ij}^2, \tag{14. 10}$$

wobei

$$k_{ij} = \sqrt{\frac{3}{2\pi}} \int_{-\infty}^{\infty} p_{i-1}(\xi)\, p_{j-1}(\xi)\, e^{-\frac{3}{2}\xi^2}\, d\xi \tag{14. 11}$$

ist.

Aus (14. 6) folgt, daß die $p_i(\xi)$ mit geradem i nur gerade Potenzen von ξ enthalten, die mit ungeradem i nur ungerade Potenzen. Deshalb sind die k_{ij}, bei denen der eine Index gerade und der andere ungerade ist, gleich 0. Folglich enthält die Reihe (14. 10) nur gerade Potenzen von r, so daß man sie auch folgendermaßen schreiben kann:

$$\Phi^2 = \frac{15}{2\pi^2} \sum_{n=1}^{\infty} K_n r^{2n}, \tag{14. 12}$$

wobei

$$K_n = \sum_{i=1}^{2n-1} \frac{k_{i,\,2n-i}^2}{i!\, (2n-i)!} \tag{14. 13}$$

ist. Da ferner $k_{ij} = k_{ji}$ ist, so hat man für $n \geqq 2$

$$K_n = 2 \sum_{i=1}^{n-1} \frac{k_{i,\,2n-i}^2}{i!\, (2n-i)!} + \frac{k_{n,\,n}^2}{(n!)^2}. \tag{14. 13a}$$

Für die k_{ij} gilt die folgende Rekursionsformel, die man durch Anwendung der Beziehungen (14. 6) und (14. 9) auf (14. 11) und partielle Integration findet:

$$k_{ij} = -\frac{2}{3}(i-2)\, k_{i-2,\,j} + \frac{1}{3}(j-1)\, k_{i-1,\,j-1}. \tag{14. 14}$$

Die Gleichung gilt für $i \geqq 2$, $j \geqq 1$, und da $k_{ij} = k_{ji}$ ist, lassen sich mit ihrer Hilfe alle k_{ij} berechnen, wenn k_{11} bekannt ist. Es ist aber

$$k_{11} = \sqrt{\frac{3}{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{3}{2}\xi^2} d\xi = 1. \tag{14. 15}$$

Berechnet man auf Grund der angegebenen Formeln die ersten Koeffizienten K_n, so erhält man für die Reihe in (14. 12)

$$\Phi^2 = 0{,}759\,91 r^2 + 0{,}133\,69 r^4 + 0{,}045\,97 r^6 + \\ + 0{,}021\,03 r^8 + 0{,}011\,35 r^{10} + 0{,}006\,84 r^{12} + 0{,}004\,46 r^{14} + \cdots. \tag{14. 16}$$

Für $r = 1$ beträgt die Summe der angeschriebenen Glieder 0,983 (statt 1,000). Diese Glieder reichen aus, um Φ^2 für $r \leqq 0{,}9$ auf zwei Dezimalstellen genau zu berechnen. Tabelle 5 zeigt die Größe von Φ für verschiedene Werte von r.

Tabelle 5.

r	Φ	r	Φ	r	Φ
0,0	0,000	0,4	0,35	0,8	0,75
0,1	0,087	0,5	0,45	0,9	0,86
0,2	0,17	0,6	0,54	1,0	1,00
0,3	0,26	0,7	0,64		

Φ ist durchweg kleiner als $|r|$, aber größer als $\frac{1}{\pi}\sqrt{\frac{15}{2}}\,|r| = 0{,}87\,|r|$. Da aus (14. 2) folgt, daß bei Gausscher Verteilung $|r| \geqq |\varrho|$ ist, so hat man auch $\Phi \geqq 0{,}87\,|\varrho|$. Das ist eine Verbesserung der Besselschen Ungleichung (12. 2), die besagt, daß bei *jeder* Verteilung $\Phi \geqq 0{,}79\,|\varrho|$ ist. Daß $\Phi < |r|$ ist, gilt natürlich nur für die Gausssche Verteilung, da im allgemeinen aus $r = 0$ noch nicht die stochastische Unabhängigkeit folgt, so daß sehr wohl $r = 0$, aber $\Phi > 0$ sein kann.

Das Maximum der Differenz $r - \Phi$ wird zwischen $r = 0{,}6$ und $r = 0{,}7$ erreicht und beträgt annähernd 0,06.

Lebenslauf.

Ich, Wassilij Höffding, wurde am 12. Juni 1914 in St. Petersburg als Sohn russischer Eltern geboren. Mein Vater Waldemar Höffding, der einer dänischen Familie entstammt, ist Volkswirt. Nach der bolschewistischen Revolution sah sich meine Familie genötigt, Rußland zu verlassen, und begab sich 1920 nach Dänemark, wo ich in eine Vorschule eintrat. 1922 siedelte meine Familie nach Hamburg um, wo ich anfangs eine Vorschule und ab Ostern 1924 eine Oberrealschule besuchte. Ende 1924 trat ich in die Leibniz-Oberrealschule in Berlin-Charlottenburg ein, an der ich Ostern 1933 die Reifeprüfung ablegte. Nachdem ich zwei Semester an der Berliner Handelshochschule studiert hatte, wurde ich am 14. Mai 1934 an der Universität Berlin immatrikuliert. Hier widmete ich mich vornehmlich dem Studium der reinen und angewandten Mathematik sowie der Volkswirtschaftslehre. Nach zehn Studiensemestern ließ ich mich Ende des Sommersemesters 1939 exmatrikulieren.